Journal of Hydrology 628 (2024) 130458

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Deep learning in hydrology and water resources disciplines: concepts, methods, applications, and research directions

Kumar Puran Tripathy [a], Ashok K. Mishra [a,b,*]

[a] School of Civil and Environmental Engineering and Earth Sciences, Clemson University, Clemson, SC, USA
[b] Zachry Department of Civil & Environmental Engineering, Texas A&M University, College Station, TX, USA

ARTICLE INFO

This manuscript was handled by marco borga, Editor-in-Chief, with the assistance of Yadu Pokhrel, Associate Editor

Keywords:
Deep learning
Hydrology
Water resources
Physics guided deep learning
Explainable Artificial Intelligence (XAI)

ABSTRACT

Over the past few years, Deep Learning (DL) methods have garnered substantial recognition within the field of hydrology and water resources applications. Beginning with a discussion on fundamental concepts of DL, we discussed the state-of-the-art DL architectures such as Long-Short-Term-Memory (LSTM), Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), and Encoder-Decoder models that have gained much attention over the recent years. The recent advancements in the DL model, such as the Attention model and Transformer Neural Network, that are designed to handle sequential time series data, are also discussed. An overview of integrating physics-based hydrological models with state-of-the-art DL models, known as Physics-Guided Deep Learning (PGDL), and its potential for improving the accuracy and interpretability of hydrological predictions are discussed. We emphasized that PGDL has the potential to enhance the physical consistency and robustness of the hydrologic predictions. We further delve into Explainable Artificial Intelligence (XAI), examining various techniques for constructing interpretable models. The objective is to empower users to comprehend and confidently trust machine learning algorithms' results (model outputs). Furthermore, we delved into the diverse applications of Deep Learning (DL) in hydrology and water resources sectors, encompassing areas such as drought and flood forecasting, remote sensing applications, water quality assessments, subsurface flow inversion problems, groundwater level prediction, and hydro-climate variable downscaling.

1. Introduction

Hydrology and water resources disciplines rely on diverse data sets and modeling techniques to study the water distribution, circulation, quantity, and quality on and below the earth's surface. Different types of models are often used to study spatio-temporal variability of hydrologic fluxes, water security (Gowri et al., 2021; Veettil et al., 2022), extreme events (Arnold et al., 1998; Liang et al., 1994; Mukherjee and Mishra, 2021; Tripathy et al., 2023), water quality (Arnold et al., 1998; Donigian et al., 1995), sediment and contaminant transport (Brunner, 1996), and climate change impact assessments (Bhatta et al., 2019; Liang et al., 1996, 1994; Niu et al., 2011; Oleson et al., 2008; Sreeparvathy and Srinivas, 2022). However, a crucial challenge lies in understanding the non-linear interactions between various processes at different spatio-temporal scales, which significantly influences the dynamics and predictability of the water cycle and extreme events (Abrahart and See, 2007; Konapala and Mishra, 2020; Sivapalan et al., 2003).

Physics (process) based hydrological models have been in use since the 1960s (Arnold et al., 1998; Beven, 1989; Crawford and Linsley, 1966; Singh and Woolhiser, 2002), for example, Soil and Water Assessment Tool (SWAT; Arnold et al., 1998), the Hydrologic Modeling System (HEC-HMS; Scharfferberg and Fleming, 2006), Joint UK Land Environment Simulator (JULES; Best et al., 2011), and Variable Infiltration Capacity (VIC; Liang et al., 1994) model. However, these models often face limitations over the adequacy of model parameterizations, data quality and uncertainty, computational constraints, complexity, and usability (Clark et al., 2017; Fatichi et al., 2016). Some of the physics-based models may not adequately capture the spatial dependence structure of the hydrologic processes occurring at different spatiotemporal scales (Blöschl and Sivapalan, 1995; Troy et al., 2008; Wood et al., 2011), and the likelihood of higher uncertainties due to discrepancies in data sets, model parameters, and model structure (Liu and Gupta, 2007; Wiens et al., 2009).

The application of data-driven models in hydrology and water resources has gained momentum in recent decades. These data-driven models are built upon statistical and machine learning techniques that

* Corresponding author.
E-mail address: ashokm@g.clemson.edu (A.K. Mishra).

https://doi.org/10.1016/j.jhydrol.2023.130458
Received 15 March 2023; Received in revised form 30 August 2023; Accepted 30 October 2023
Available online 15 November 2023

learn patterns directly from the observed data. Examples include Artificial Neural Networks (ANN; Maier and Dandy, 1996), Support Vector Machines (SVM; Drucker et al., 1999), and Random Forest algorithms (Breiman, 2001). The data-driven methods used in model calibration, parameter estimation, and handling various sources of uncertainties are perceived as less challenging compared to the physics-based models (Ghaith et al., 2020; Liu et al., 2021; Sikorska et al., 2015).

The recent surge in hydrological, geological, and climate science-related data sets' volume, variety, and veracity provides new opportunities (Reichstein et al., 2019) for studying hydro-climatic processes and real-world water resources applications. Hydrological data typically emanates from multiple sources, such as observed and reanalysis databases (e.g., ERA5; Hersbach et al., 2020), remote sensing products (e.g., Landsat, MODIS, and soil moisture missions; Entekhabi et al., 2010), climate models (e.g., Coupled Model Intercomparison Project Phase 6 (CMIP6); Weather Research and Forecasting Model, Skamarock et al., 2008), sensor data (Wang et al., 2008), camera rain gauges (Allamano et al., 2015), unmanned aerial photogrammetry, Internet of Things (IoT) measurements (Oguz et al., 2022), and crowdsourcing data sets (Sermet et al., 2020). The key challenge in dealing with these heterogeneous data sets is primarily associated with the difference in spatial and temporal resolution as well as the volume of the data sets. For example, the total volume of the CMIP-6 data archive is estimated to be approximately 15–30 petabytes (Reichstein et al., 2019). The main challenge is extracting and detecting patterns of hidden information from the data deluge, also known as the data discoverability problem (Máchová et al., 2018). Therefore, extracting useful information from high-dimensional and multi-scale hydrological datasets will provide new avenues for quantifying hidden processes, thereby advancing hydroclimatic prediction, detection, and attribution.

In the past decade, data-driven models such as Machine Learning (ML) and Deep Learning (DL) methods have garnered interest in hydrology and water resources communities. Commonly used data driven models are neural network models (Adamowski and Chan, 2011; Amaranto et al., 2019; Elshorbagy et al., 2010; Guzman et al., 2017; Sun et al., 2022), self-organized map (SOM; Chang et al., 2021; Nourani et al., 2013), boosted regression trees (Rosecrans et al., 2017), multivariate adaptive regression splines and M5 model trees (Rezaie-balf et al., 2017), random forest (RF; Konapala and Mishra, 2020; Schoppa et al., 2020; Yu et al., 2017) and support vector machines (SVM; Liu et al., 2021; Raghavendra. N and Deka, 2014; Shabri and Suhartono, 2012).

The DL methods are instrumental in addressing challenging problems in the field of hydrology and water resources, such as time-series forecasting (Kao et al., 2020; Yang et al., 2019a), land use and land cover (LULC) classification (Maggiori et al., 2017; Zhang et al., 2020), hydrodynamic modeling, downscaling (Wang et al., 2021), change and anomaly detection (Zhong et al., 2019), hurricane tracking (Kim et al., 2019), water quality assessment (Prasad et al., 2022), extreme weather prediction (Zhou et al., 2019), generation of complex multimodal data distributions (Laloy et al., 2018). DL models' performance can be superior to the state-of-the-art process-based models and traditional ML approaches, especially in data-rich environments (e.g., large volumes and high-quality data). The DL models have improved the multi-scale (Fang et al., 2017), multi-task (Sadler et al., 2022), and multimodality barrier with high-fidelity solutions (Laloy et al., 2018). One of the key advantages of DL is its ability to ease the manual task of feature engineering. However, it's important to note that the effectiveness of this automation is highly dependent on the specific use case and data type, and careful manual feature engineering is still essential in many applications. Feature engineering is the process of transforming raw data into suitable representations or features using mathematical functions. DL enables the detection of complex nonlinear relationships between inputs and outputs through successive, deeper layers of feature engineering. Several factors complemented the emergence of DL in hydrology and water resource disciplines, which includes: (1) availability of large volumes of data, (2) rapid progress in the parallel computing machines with multi-core options, graphical processing units (GPU), and multi-threaded execution, (3) Niche software platforms including TensorFlow (Abadi et al., 2016), Keras (Chollet, 2021), PyTorch (Paszke et al., 2017), BigDL (Dai et al., 2019), Theano (The Theano Development Team et al., 2016) and Caffe (Jia et al., 2014) that allow building hierarchical DL architectures without exploring much complex mathematical details, (4) robust optimization performance achieving near-optimal solutions, and (5) improved regularization methods to overcome overfitting.

DL methods are critical to the Fourth Industrial Revolution (4IR or Industry 4.0), particularly in big data analysis (Oosthuizen, 2022). The shift towards data-driven models and data-intensive research in hydrology and water resources disciplines offers immense potential in the coming years. Although the application of DL concepts and tools has gained momentum, a comprehensive discussion of DL concepts, methodologies, applications, challenges, research gaps, and potential opportunities in hydrology and water resources disciplines is still lacking, which is the fundamental motivation for this review.

Our study outlines the DL concepts and methodologies (sections 2 and 3) and summarizes the progress of DL applications in hydrology and water science (section 5). Our review compliments recent review papers (Camps-Valls et al., 2021; Shen, 2018; Shen et al., 2018; Sit et al., 2020; Yaseen et al., 2019) by emphasizing the novel facets of DL models, such as Physics-Guided Deep Learning (PGDL) and Explainable Artificial Intelligence (XAI; Section 4), which were not discussed in previous studies. We also highlighted the advantages of the latest DL architectures, such as Transformers and Attention models, which outperform state-of-the-art LSTM networks and efficiently use parallel computing abilities that were overlooked in previous reviews. We examined the challenges and opportunities that DL provides for hydrology and water resources disciplines provided in section 6. Additionally, we delved into the unexplored potential of Physics-Guided Deep Learning and discussed the importance of Explainable Artificial Intelligence XAI.

2. DL concepts

2.1. Background of DL

Artificial Intelligence (AI) was first introduced in the 1950s to perform intellectual tasks commonly performed by humans (Dick, 2019). ML, a branch of AI draws its roots from mathematical statistics, which require a vast amount of training data to continuously discover underlying patterns, which may be linear or nonlinear. ML models possess numerous parameters that aim to capture the nonlinear patterns in the data distribution (Chollet, 2021).

DL, a subfield of ML, is a powerful method for mapping input data to target outputs using successive layers of nonlinear transformations, typically learned through neural networks (Chollet, 2021). Unlike traditional ML techniques such as Support Vector Machines (SVM; Vapnik, 1999) and Random Forests (RF; Breiman, 2001), which employ simple transformations and limited representations of the data, DL leverages multiple layers of representation to learn complex feature mappings from the input data automatically. The term "deep" in Deep Learning refers to this multi-layer architecture, and it can discover complex, nonlinear relationships in data. DL algorithms have become crucial in various applications across diverse scientific disciplines.

DL methods can be broadly classified into supervised, unsupervised, and self-supervised learning (Kotsiantis et al., 2006) (Fig. 1). In supervised learning, essential features or rules that map the input data to desired targets are generated. When new testing data is provided, it produces the output based on the learned features. Apart from classification or regression tasks, some canonical examples of supervised learning include single-step/multi-step time-series prediction, object detection, and image segmentation. Unlike supervised learning, unsupervised learning analyzes and clusters unlabeled datasets and discovers hidden patterns or data groupings to discover similarities and

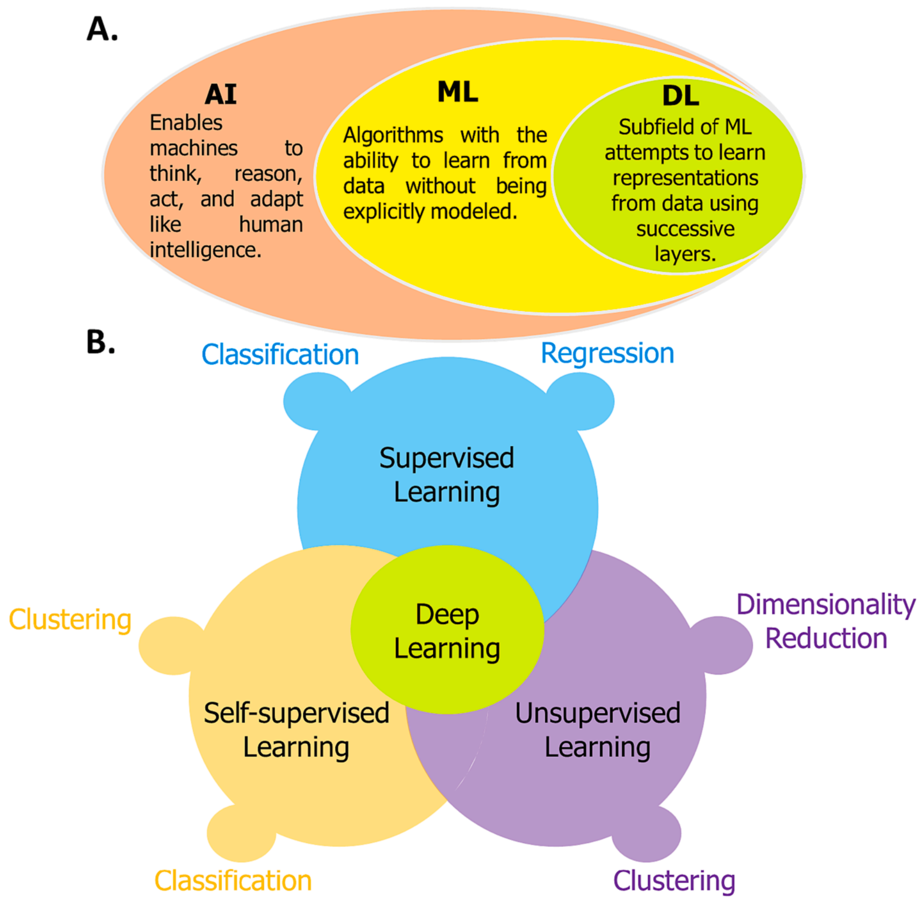

Fig. 1. (A) Illustrates the hierarchical relationship between Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL), with AI representing a broader field, ML as its subset, and DL as a subset of machine learning, and (B) Details the three core components of DL: supervised learning for classification and regression tasks, self-supervised learning for clustering and classification, and unsupervised learning for clustering and dimensionality reduction.

differences in information, making it an ideal solution for exploratory data analysis (Hastie et al., 2009). This form of machine learning attempts to better understand the structure and distribution of the input dataset by data visualization, denoising, data compression, or the association among variables before applying supervised learning. K-mean clustering (MacQueen, 1967), Principal Component Analysis (PCA; Pearson, 1901), and Generative Adversarial Networks (GAN; section 3.4) are popular unsupervised techniques. Self-supervised learning is a special type of supervised learning where there are pseudo-labels, unlike labels or ground truths in supervised learning. These pseudo-labels are usually generated from the input data using a heuristic algorithm. Autoencoders (section 3.3) are well-known examples of self-supervised learning, where the DL model learns to compress and encode data so that the reduced encoded information can be reconstructed as close to the original input as possible. There is another branch of machine learning called Reinforcement Learning (RL; Sutton and Barto, 2018) which slightly varies from DL. In DL, the idea is to decode the features/rules that map a set of inputs to certain outputs; in contrast, in RL, an intelligent agent continuously learns by trial and error procedure in an interactive environment using feedback from its own actions and experiences. However, it's important to mention that DL and RL can overlap, such as in Deep Reinforcement Learning. This approach integrates DL methods within an RL framework, creating a powerful tool to enhance learning and predictive capabilities.

2.2. How does DL work?

DL works by mapping inputs to outputs through multiple layers that perform various mathematical transformations. The model's " parameters, " including weights and biases, are iteratively adjusted to optimize this mapping - a process known as learning (Chollet, 2021; Gulli and Pal, 2017).

In a DL network, hidden layers apply nonlinear transformations on inputs using activation functions (Fig. 2). After an output is predicted, a loss function computes the difference between this prediction and the actual output, indicating the model's accuracy. This error is then used to adjust the weights and minimize future errors through backpropagation, an optimization algorithm (Rumelhart et al., 1986; Fig. 2). For a simple neural network model demonstration, the reader can follow supplementary section A1 for the backpropagation algorithm concept. Starting with random weights, iterations refine these values to reduce the loss function and accurately predict the target. Further information on key components like loss functions, activation functions, and optimizers is available in supplementary sections A2 and A3.

2.3. Paradox of Underfitting-Overfitting issues

Fig. 3a illustrates underfitting, overfitting, and optimal fitting in a regression model. Underfitting, indicative of a too simplistic model, fails to capture the relationship between variables, causing high bias and underperformance on all data. Overfitting, conversely, happens when an overly complex model incorporates noise from training data, causing high variance and suboptimal test data performance. Optimal fitting strikes a balance between complexity and simplicity, accurately representing variable relationships without fitting to noise, thus ensuring good performance on training and test data.

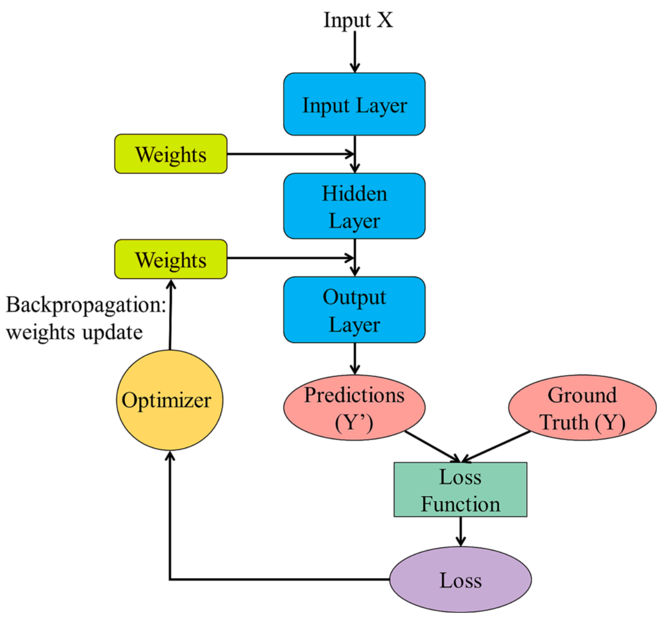

Fig. 2. Schematic representation of the training process in a DL model. The diagram demonstrates the forward pass through the hidden layers, the calculation of loss using the loss function, and the iterative refinement of model parameters via the backpropagation algorithm.

2.4. *Regularization*

Regularization is a crucial aspect of DL model training, balancing optimization (adjusting model weights to achieve the best results on the training set) with generalization (the model's performance on unseen data). A DL algorithm aims to maximize this generalization capability (Goodfellow et al., 2016). At the onset of training, the model learns data patterns, improving its performance as the number of epochs increases. However, further optimization doesn't enhance generalization beyond a certain point, leading to overfitting as the model begins learning patterns exclusive to the training data that are misleading when applied to new data (Kukačka et al., 2017). Unfortunately, the model's generalization capability can not be controlled directly; instead, we opt for various regularization methods to improve the model's generalization capability. The regularization techniques are one of the critical steps for developing architectures of the DL models. We briefly discuss the commonly used regularization techniques.

2.4.1. *Weight regularization*

The concept of weight regularization is to adjust the weights by updating the learning algorithm to reduce overfitting and improve the model's generalization. The overfitting problems are likely to be observed in complex models compared to simple models (fewer parameters and/or lesser variability in parameters). We penalize the higher weight values in weight regularization by adding a cost to their loss function (Tian and Zhang, 2022). For example, in the least Absolute Shrinkage and Selection Operator (LASSO; Tibshirani, 1996), linear regression model $Y = X^T\beta$, where Y is the response variable, X is the predictor set and β is the vector of linear regression coefficients, we seek to minimize the cost function $\sum_{i=1}^{n}\left[Y - X^T\beta\right]^2 + \lambda\sum_{j=1}^{p}|\beta|$, where λ is the regularization parameter and $\lambda\sum_{j=1}^{p}|\beta|$ is the penalty term.

The weight regularization in a DL task is performed either via L1 or L2 regularization methods. In the L1 regularization method (Tibshirani, 1996), the cost added is proportional to the absolute value of weight coefficients. This is analogous to the LASSO linear regression, where we seek to minimize the following:

$$\widehat{\beta} = \frac{argmin}{\beta}\left[\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right|\right] \tag{1}$$

In L2 regularization method (Hoerl and Kennard, 1970), the cost added is proportional to the square of the value of weight coefficients, which is analogous to the Ridge regression as provided below:

$$\widehat{\beta} = \frac{argmin}{\beta}\left[\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2\right] \tag{2}$$

The penalized term improves the model performance by addressing overfitting.

2.4.2. *Dropout regularization*

Dropout regularization is widely used to prevent overfitting problems in DL models (Srivastava et al., 2014). It works by randomly deactivating a fraction of the nodes in a layer during each training epoch. This randomness ensures no unit is entirely dependent on another, boosting the model's generalization capability by breaking co-adaptations among nodes (Hinton et al., 2012). In other words, dropout prevents overfitting by decreasing the dependence between the nodes in the hidden layers. Essentially, dropout results in an ensemble of smaller, less biased subnetworks that yield more robust predictions (Baldi and Sadowski, 2013).

However, using dropout requires caution. Applying it just before the output layer can degrade performance, as the network doesn't get the chance to correct errors adequately. Similarly, in small networks, dropout can excessively reduce the network size, affecting the learning process. Lastly, dropout before Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) layers can hinder proper back-propagation due to the random zeroing of temporal units (Gal and Ghahramani, 2016).

2.4.3. *Early stopping*

Identifying the stopping time (number of epochs) for training the DL model is often unclear. Too many epochs lead to overfitting, and too little results in underfitting of the models. Early stopping criteria (Fig. 3c) ensure that the training should be stopped when the validation error starts building up after a certain number of epochs, indicating no improvement of generalization (Yao et al., 2007). Stopping the training before the generalization error increases (when the validation loss is the least) will help reduce overfitting.

2.4.4. *Batch normalization*

Batch normalization is a technique (see Fig. 4a) used to optimize the training of deep neural networks by addressing the problem of internal covariate shifts (Ioffe and Szegedy, 2015). Internal covariate shift refers to the change in the input data distribution to subsequent layers during the network's training process due to frequent weight adjustments, which can destabilize learning.

Batch normalization optimizes deep network training by standardizing each mini-batch's input data, maintaining a steady distribution, and enhancing efficiency. It computes each mini-batch's mean and standard deviation for normalization and incorporates learned scale and shift parameters for flexible training. Note that batch normalization should not be used with dropout as the random deactivation of neurons can interfere with the standardization process (Goodfellow et al., 2016).

2.4.5. *k-fold cross-validation*

K-fold cross-validation (Fig. 4b) is an effective resampling approach for evaluating the DL models when limited training data are available (Gulli and Pal, 2017). This procedure randomly splits the data sample into k number of groups; for example, when k = 5, it is called 5-fold cross-validation. This procedure helps estimate the skills of a DL model for unseen data not used during the model's training. The data are

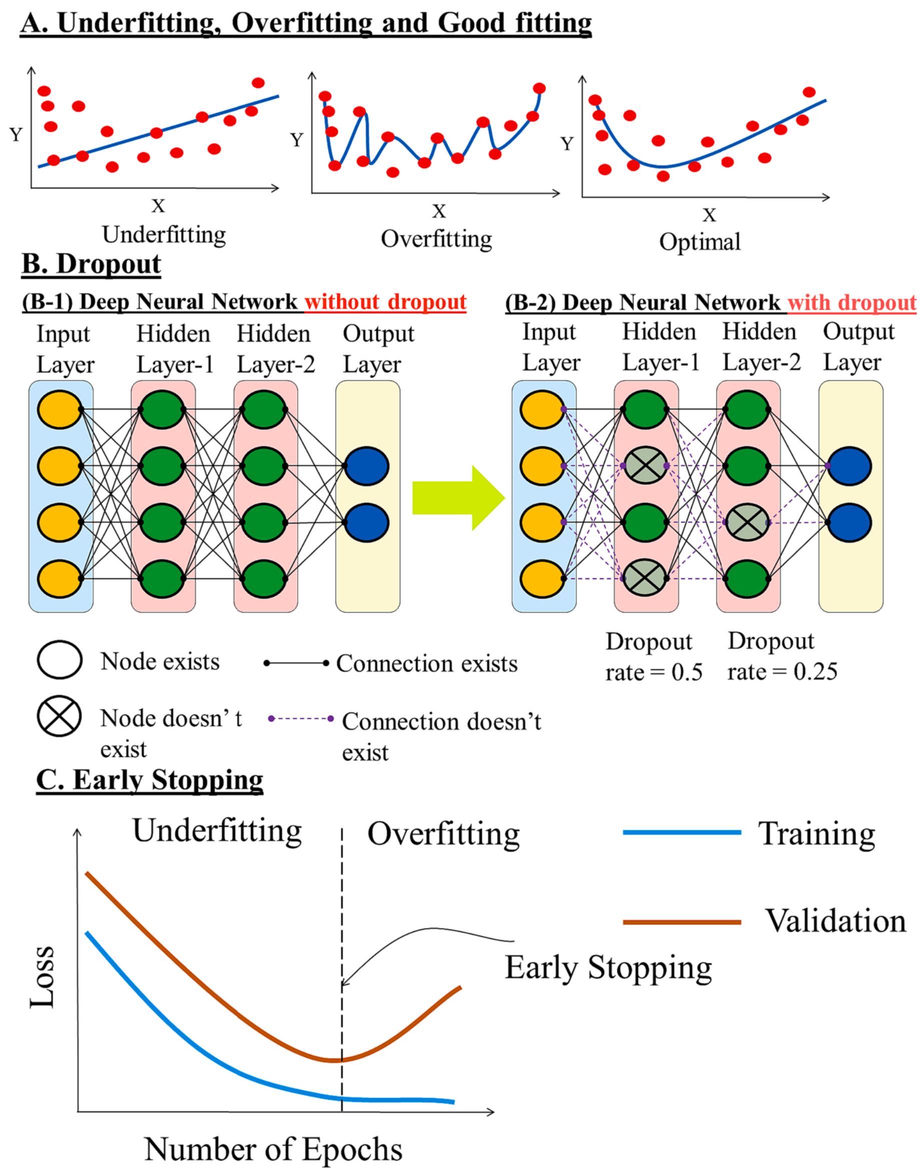

Fig. 3. (A) Illustrating the concepts of underfitting, overfitting, and optimal fitting between independent and dependent variable data. (B) Dropout regularization is applied to the DNN (B-2) with dropout rates of 0.5 and 0.25 to the hidden layer-1 and −2. (C) Early stopping by monitoring the validation loss during training and stopping the training process when the validation loss increases after a certain number of epochs.

randomly split into training and validation sets. The model performance is then averaged over all k iterations to get a more robust estimate of the model's performance on new, unseen data.

2.4.6. Data augmentation

Data augmentation enhances datasets by generating new training samples from existing data (Simard et al., 2003). In image-related applications, this involves transformations like flipping or rotating images, which expand the data's diversity and help counter overfitting (Goodfellow et al., 2016). This exposure to varied data can improve the model's generalization, making it more robust (Krizhevsky et al., 2012). For time series data, augmentation methods such as shifting or adding noise can bolster the model's resilience to data variation, like changing trends or seasonal patterns (Iwana and Uchida, 2021; Wen et al., 2021).

3. Deep learning architectures

The following section discusses DL architectures commonly used in hydrology and water resources applications. A quick overview and comparison of these architectures in terms of their specific uses in hydrology and water resources disciplines, along with their advantages and disadvantages, are provided in Table 1.

3.1. Convolutional neural network (CNN)

Convolutional Neural Networks (CNN; Krizhevsky et al., 2012; Lecun et al., 1998) are a special kind of DL model suitable for analyzing image-based datasets and grid-based cross-sectional time-series data due to their ability to capture spatial structure (relation between neighboring pixels) and local connectivity between the pixels. This ability to capture spatial relationships in hydrology is advantageous when dealing with geospatial datasets such as satellite imagery or digital elevation models (DEMs), making CNNs a powerful tool for hydrological modeling. Over the years, CNNs have seen remarkable advancements with powerful architectures such as AlexNet (Krizhevsky et al., 2012), VGGNet (Simonyan and Zisserman, 2015), GoogLeNet (Szegedy et al., 2014), and ResNet (He et al., 2016), which have made significant strides for image classification including land cover classification, identification of water

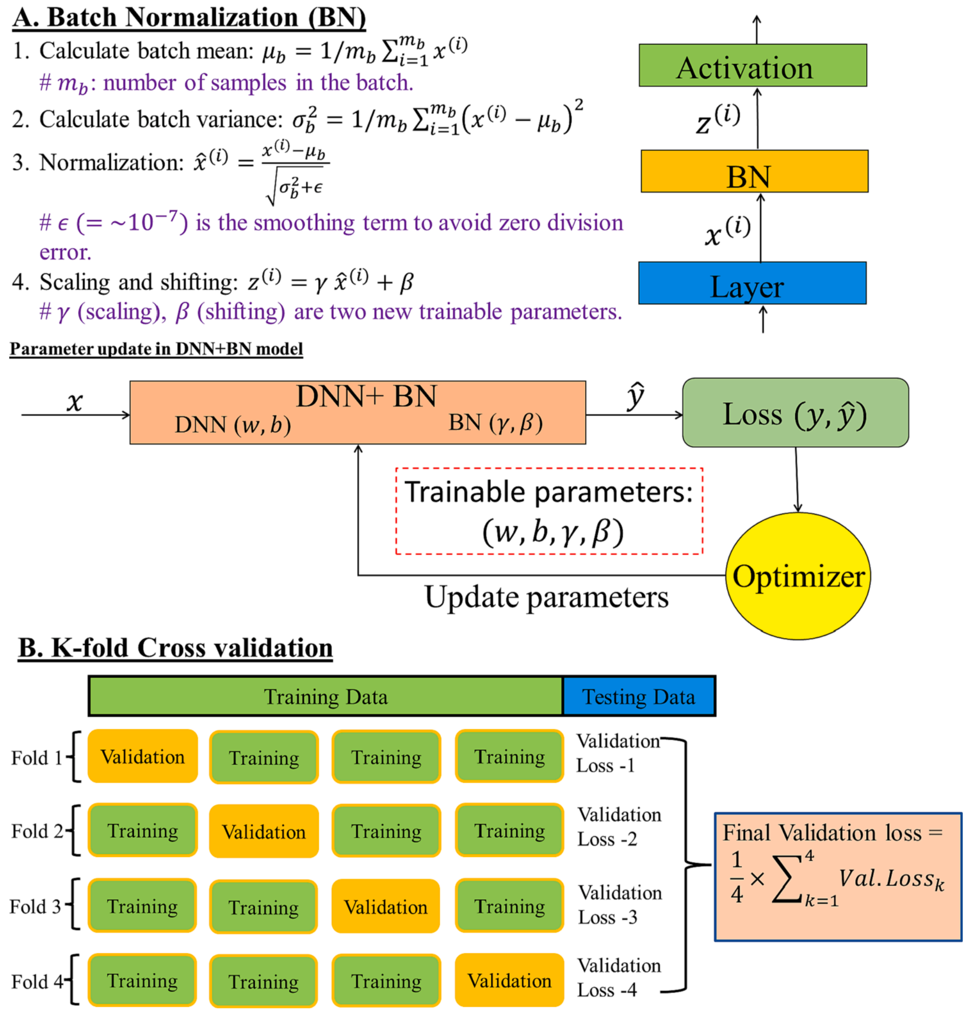

Fig. 4. (A) Batch Normalization is applied after a layer in a deep neural network (DNN + BN). BN introduces two additional trainable parameters: scale (γ) and shift (β) along with the weights and biases for each neuron in the layer. (Bottom panel) All the trainable parameters are updated using the backpropagation algorithm in each iteration using an optimizer such as Adam optimizer. (B) The K-fold Cross-Validation Process: The training data is divided into distinct sets (folds), with each subset serving as a validation set in turn, while the remaining data is used for training. The final validation score is the mean of the validation losses across all folds, offering a more robust performance estimate on unseen data.

bodies, and forecasting of rainfall events based on cloud imagery.

The major disadvantage of early-generation Multi-Layer Perceptron (MLP) is that the spatial information is lost by flattening the 2-D image input data to 1-D. In contrast, CNNs retain spatial information by employing a linear mathematical operation called *convolution*. Fig. 5 illustrates a typical CNN model for an image classification task. Typically, a CNN includes three types of layers: convolution layers, and pooling layers, followed by an MLP or fully connected neural network (FCNN).

The convolution operation involves traversing 2D convolution kernels (filters) to detect the critical features/patterns from the input image (See Fig. 5b). The parameters of these filters are updated iteratively using the backpropagation algorithm. The features extracted after the convolutional layer are subsampled by a pooling layer. For example, a two-by-two max-pooling layer will select the maximum values from each two-by-two neighboring window (Fig. 5b), which shrinks the input feature map by 75 %. A real-world image classification task involves multiple convolutional filters (such as 32 or 64) that transform original input images by a series of convolution and pooling layers, which helps to capture high-level feature maps in the input images. This feature extraction process is highly relevant in hydrological studies where patterns and anomalies in water bodies, land use changes, or soil moisture distribution can be effectively detected. These extracted feature maps thus have much fewer parameters than the original input image, allowing the model to learn more quickly than the MLPs. Finally, an MLP or FCNN model is implemented to classify the images based on the extracted feature maps from the convolution and pooling operations series. CNN can also be used for hydrologic time series prediction/forecasting. Just like 2D-CNNs capture the spatial dependence by extracting features from local input 2D patches, their 1D counterparts, called 1D-CNN, can recognize the local patterns in time series or sequential problems.

3.2. Long-Short term Memory (LSTM) networks

Recurrent Neural Networks (RNNs) are proficient in processing time-series data, an important aspect when predicting hydrological variables like streamflow, groundwater level, or rainfall intensity. However, standard RNNs struggle to handle large volumes of contextual information due to the “vanishing gradient problem” (Hochreiter, 1998). The LSTM model, a specialized form of RNN, was introduced to overcome this issue. It emphasizes preserving relevant past information while disregarding irrelevant data (Hochreiter and Schmidhuber, 1997).

The LSTM network depicted in Fig. 6a is comprised of a sequence of LSTM cells followed by a fully connected neural network (FCNN) layer. The network processes input data at each time step, with each LSTM cell

Table 1
Comparison of Deep Learning Models for Hydrological Applications.

DL Model	Examples of Hydrologic Applications	Advantages	Disadvantages
Convolutional Neural Network (CNN)	Analyzing spatial patterns in hydrological data (e.g., precipitation, soil moisture maps), Land-use/Land-cover (LULC) classification from satellite images	Effectively captures spatial dependencies and requires fewer parameters, thus limiting overfitting.	Potential challenge in dealing with heterogenous data that doesn't represent grid-like topology.
Long Short-Term Memory (LSTM)	Time-series forecasting of hydrological variables (streamflow, groundwater levels), Weather forecasting	It efficiently captures temporal dependencies and is robust to the vanishing gradient problem.	Computationally intensive, model interpretability is limited.
Autoencoders	Anomaly detection in hydrological data, Extraction of important features from high-dimensional data (e.g., multispectral images)	Efficiently learns data codings, useful for unsupervised learning tasks	It can be sensitive to the input data and might reproduce noise.
Generative Adversarial Networks (GANs)	Generation of synthetic yet realistic hydrological data, Downscaling of climate variables	Capable of generating new data with the same statistics as the training set	Training can be complex and requires a balance between the generator and discriminator.
Encoder-Decoder model	Multi-step ahead prediction in hydrological modeling, Data assimilation	Separation of encoding and decoding tasks enables better performance	Requires paired examples for training, may struggle with very long input sequences of time series data
Attention Model	Handling seq2seq prediction tasks with long-range dependencies (e.g., seasonal precipitation forecasting)	Focuses on specific parts of the input sequence, which can improve results for specific tasks	Increases computational complexity
Transformer	Multivariate hydrological forecasting (e.g., simultaneous prediction of rainfall, streamflow, and evapotranspiration)	Achieves high performance with parallelization, efficient handling of long sequences	Requires higher volume of training data and more computationally intensive than RNNs

generating an output. These outputs are then collectively processed by the FCNN layer to generate the final prediction. Fig. 6b demonstrates an LSTM cell at timestep t. Its core is the cell state (C_t), regulated by three gates: input, output, and forget gate. The forget gate (f_t) decides whether to retain or discard the previous time step's information using a *sigmoid* activation by examining the previous hidden state output (h_{t-1}) and input (x_t).

$$\mathrm{f_t} = \mathrm{sigmoid}\left(\mathrm{w_f}[\mathrm{h_{t-1}}, \mathrm{X_t}] + \mathrm{b_f}\right) \tag{3}$$

where, *w* and *b* are the weights and biases.

The input gate determines what new information content can be added to the cell state (C_t). It has two components: a *sigmoid* layer and a *tanh* layer. The *sigmoid* layer decides which values to update, and the *tanh* layer filters whether the new information is relevant to the context of time series prediction; if it's relevant, it gets updated to the cell state; otherwise, it is removed. Both *sigmoid* and *tanh* layers update new information to the cell state through pointwise multiplication. This mechanism of updating new information based on relevance is crucial for hydrological modeling as it helps in capturing only significant patterns from past data.

$$i_t = sigmoid(w_i[h_{t-1}, X_t] + b_i) \tag{4}$$

$$\widetilde{C}_t = tanh(w_c[h_{t-1}, X_t] + b_C) \tag{5}$$

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C}_t \tag{6}$$

The output gate of an LSTM cell determines the information to be transferred to the next cell and the final output. Using a sigmoid layer and a *tanh* layer, it filters and transmits only the relevant information from the current cell state ($\mathrm{C_t}$) to the subsequent time step $\mathrm{t}+1$ through a *tanh* layer.

$$o_t = sigmoid(w_o[h_{t-1}, X_t] + b_o) \tag{7}$$

$$h_t = o_t \times tanh(C_t) \tag{8}$$

The output of the cell at time *t* is a function of the long-term memory (cell state C_t) and the short-term memory (current hidden state h_t), allowing the network to remember both long-term and short-term information. This is particularly beneficial in hydrological scenarios where both short-term (e.g., recent precipitation events) and long-term (e.g., seasonal variations) factors play a role in the outcomes (Jiang et al., 2022).

An extension to the standard LSTM is the Bidirectional LSTM (BiLSTM; Graves and Schmidhuber, 2005). BiLSTM incorporates two LSTM layers, one for positive time direction (forward state) and another for negative time direction (backward state). This architecture allows the model to have access to both past (backward) and future (forward) contexts, which can lead to improved performance on various tasks by capturing the patterns that depend on the context of the input sequence (Graves and Schmidhuber, 2005). For instance, understanding past precipitation events and future weather forecasts in flood prediction is vital, making BiLSTM highly advantageous.

There are other variants of the GRU also exist. The Gated Feedback Recurrent Network (GRFN; Chung et al., 2014) is an example where gating mechanisms are used to control the flow of information in the hidden state of recurrent neural networks, allowing for better control over the learning process. The GRFN improves upon the GRU by adding a gating unit that modulates the hidden-to-hidden recurrent connections, thereby providing additional complexity in learning the temporal dependencies.

3.3. Autoencoders (AE)

Autoencoders (See Fig. 7), a type of self-supervised network, are relevant in hydrology for feature extraction and data dimensionality reduction (Jiang, 2018). They typically consist of an encoder, a bottleneck, and a decoder. The input data are compressed via an encoder to form a bottleneck, and then the decoder reconstructs the input using this reduced feature space (Ballard, 1987; Rumelhart et al., 1985).

Autoencoders are self-supervised, as their training assigns their own inputs as the output targets. They create an information bottleneck by restricting the hidden layer dimension to smaller than the input, leading the encoder to learn salient data features, which the decoder uses for reconstruction. In hydrology, autoencoders' ability to capture key features from input data can be critical, assisting in predicting future hydrologic events or understanding processes.

The training of autoencoders involves iteratively minimizing the loss function and updating the network weights. After training, only the encoder block is used to generate a low-dimensional representation of the input data. Autoencoders have an advantage over PCA, capturing both linear and non-linear patterns present in the training data (Goodfellow et al., 2016; Hinton and Salakhutdinov, 2006), useful when dealing with complex, non-linear hydrological systems.

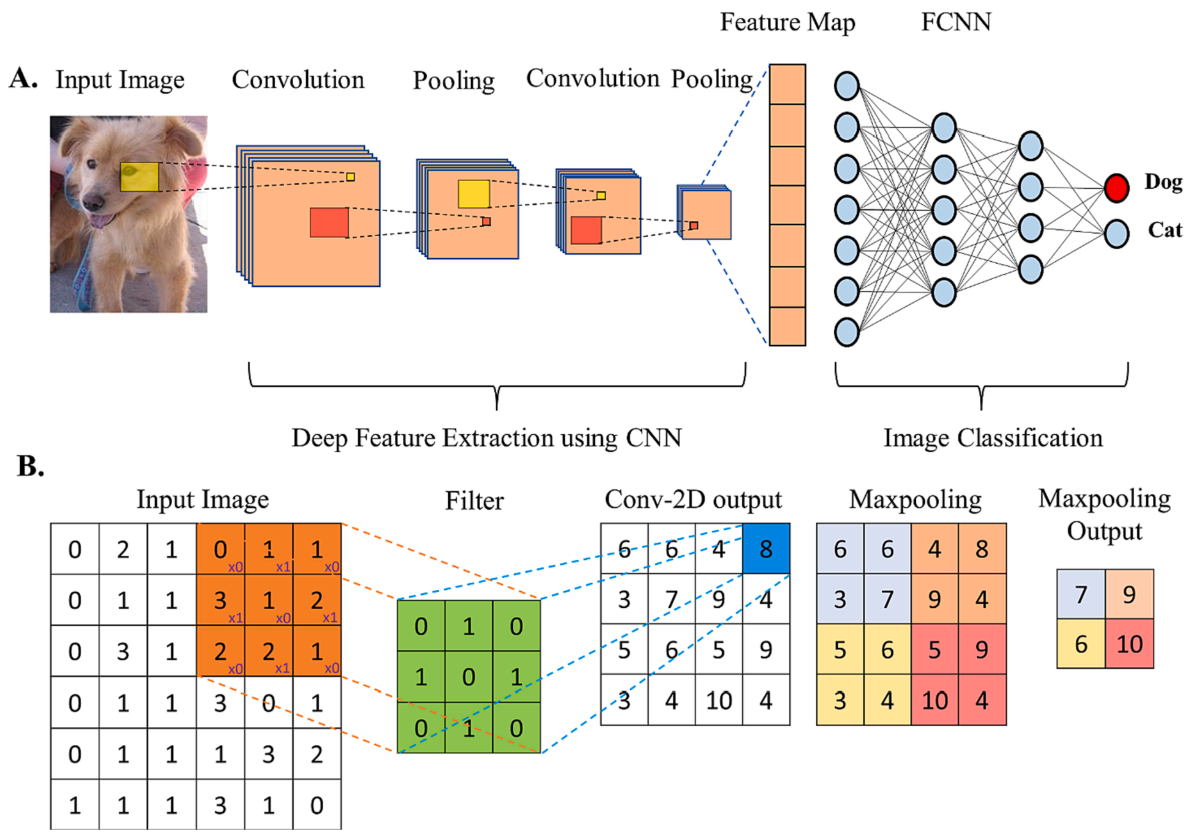

Fig. 5. A schematic representation of a CNN architecture for binary classification. The architecture comprises three key components: convolutional layers, pooling layers, and fully connected neural networks (FCNN).

Depending on the modeling task, either MLP, CNN, or LSTM layers can be used for the encoder and decoder parts. For instance, in hydrology, Denoising autoencoders (DAE) can be used for remote sensing applications like reconstructing or denoising satellite imagery related to water bodies or land use.

3.4. Generative Adversarial networks (GAN)

Generative Adversarial Networks (GANs) are unsupervised DL models that provides an ability to learn deep representations without extensively annotated training data (Creswell et al., 2018), which can be useful in hydrological applications where comprehensive and accurately labeled data is scarce. GANs used a clever approach by framing the problem using two sub-models (a) the generator model and (b) the discriminator model (Goodfellow et al., 2014), working in opposition to each other to learn the underlying distribution of the data.

During the training of GAN, the Generator (G) uses random input noisy data (z) from an arbitrary distribution p(z) to generate synthetic (fake) data (G(z)), which is further fed to the discriminator module (Fig. 8). The discriminator has access to the real data (X) and the synthetically generated data G(z) from the generator. The goal of the generator model is to generate samples as close as possible to real data (x), while the purpose of the discriminator is to detect whether the generated sample from the generator is real or fake, thus the name *adversarial network* (Goodfellow et al., 2014). The outputs from the discriminator model vary between 0 and 1 and they are derived using a sigmoid activation function. The loss functions of the generator and the discriminator as given by (Goodfellow et al., 2014):

$$L_{\mathrm{Discriminator}} = -E_X[\log(D(x))] - E_Z[\log(1 - D(G(z)))] \tag{9}$$

$$L_{\mathrm{Generator}} = -E_Z[\log(D(G(z))] \tag{10}$$

The discriminator tries to minimize the loss function either by maximizing D(x) or by minimizing the D(G(z)), whereas the generator attempts to maximize the D(G(z)). Remember that the function D(.) gives a real number between 0 and 1 since it has a sigmoid activation function. The training is said to be converged when the discriminator can no longer detect fake images produced by the generator and outputs a number close to 0.5 every time.

Mirza and Osindero (2014) proposed conditional GANs that learn to generate synthetic samples from the conditional distribution $p(z|y)$ instead of the marginal distribution p(z). In this approach of generative modeling, both the generator and discriminator are fed with some auxiliary feature labels (y), which help the generator produce samples consistent with the auxiliary information. This could be a valuable tool in hydrology where certain conditions or features, such as precipitation or temperature, could be used as conditional information to generate synthetic data for different scenarios.

In addition to Conditional GANs, other GAN variants have been formulated to optimize the original GAN model. One such example is Deep Convolutional GANs (DCGANs), which apply deep CNNs within the GANs framework to enhance their potential to generate high-quality images by learning hierarchical data representations (Radford et al., 2016). Another variant is Wasserstein GANs (WGANs), which introduces a new loss function based on the Wasserstein distance. This improves GANs' training stability, mitigates mode collapse issues, and provides meaningful learning for model training and evaluation (Arjovsky et al., 2017).

3.5. Encoder-Decoder model

The encoder-decoder network (Sutskever et al., 2014) provides a way to train RNNs for sequence-to-sequence (seq2seq) prediction problems (Cho et al., 2014). The encoder-decoder model (Fig. 9) consists

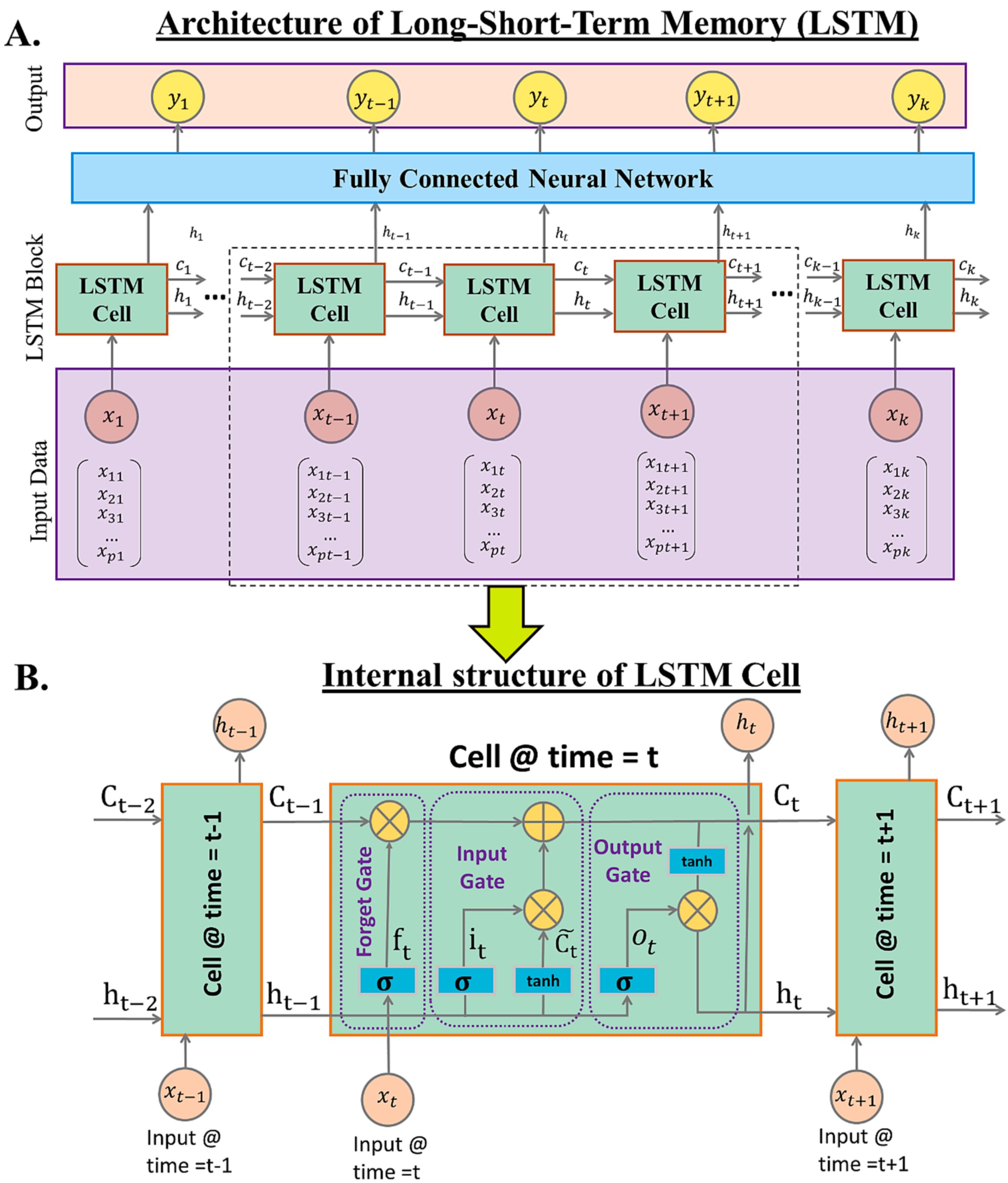

Fig. 6. The architecture of an LSTM model at time step t consists of three gates: forget gate, input gate, and output gate.

of two RNNs: an encoder and a decoder. The encoder, usually built by stacking RNNs, encodes the input sequences and produces a fixed-length context (feature) vector, a simple function of the final hidden state. The decoder RNN receives the context vector to generate the output sequences. In a seq2seq model, these two RNNs are trained jointly to maximize the conditional likelihood of generating output sequences given the input sequence training datasets. The encoder-decoder model is effective for hydrological forecasting, where the encoder processes past historical record (e.g., rainfall) to predict future conditions (e.g., river discharge). Sometimes, the encoder-decoder framework can be modified by stacking multiple LSTMs to enhance the model's accuracy by preserving contextual information over long periods (Asadi and Safabakhsh, 2020; Du et al., 2020).

The major drawback of the encoder-decoder model is when the dimension of the context vector from the encoder RNN is too small to encode long input sequences. Bahdanau et al. (2016) proposed "Attention" model (Section 3.6), which overcomes the problems of long-term dependency of the RNN-based encoder-decoder frameworks. Although encoder-decoder architectures were initially developed for seq2seq modeling tasks using RNNs, recent studies have stacked CNNs in the encoder-decoder models for extracting spatial representations for object detection and image classification tasks.

3.6. Attention model

In the domain of hydrological forecasting, sequence-based modelling plays a crucial role. The attention model, a recent development in DL, has become a significant focus seq2seq based time series forecasting,

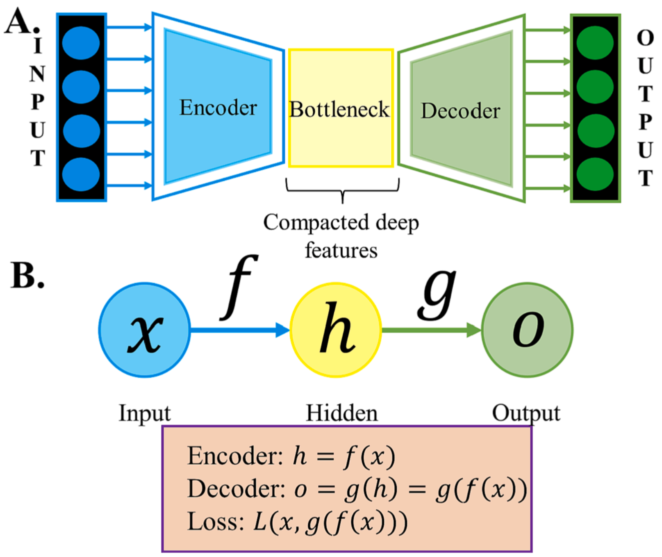

Fig. 7. Schematic diagram depicts an autoencoder architecture. Fig. 7 (A) provides a more detailed view of the autoencoder, highlighting the encoder, decoder, and a 'bottleneck' component, instrumental in capturing and compressing salient features from the input for efficient data reconstruction. Fig. 7 (B) provides the process of transforming an input (x) via an encoder function to a hidden representation (h) followed by the reconstruction to output (o) through a decoder function. The discrepancy between the original input and the output forms the loss.

particularly when modeling long-term hydrological sequences (Bahdanau et al., 2016). The key strength of attention models lies in their ability to focus on specific parts of the input data sequence when generating an output, thereby improving the model's ability to handle longer input data sequences and their complex dependencies. Unlike LSTM and encoder-decoder-based RNN models, which assign equal weights to all past lags, attention models assign more weightage to the most critical past lag information, thus enhancing efficiency and reducing computational time.

In the encoder-decoder models, the context vector is a simple function of the last hidden state of the encoder RNN, and their accuracy declines with more extended input time-series data. Bahdanau et al. (2016) improved the encoder-decoder models with slight modification by considering feeding all the hidden states of the encoder with different weights of their importance to the context vector instead of taking the last hidden state only. So, every time the decoder generates a sequence, it searches for a subset of hidden states from the encoder RNN that maximizes the conditional likelihood of generating an output sequence given the input sequence. This mechanism is called *Attention*.

The attention model, shown in Fig. 10, assigns different weights to hidden states based on their importance in the output sequence generation (Bahdanau et al., 2016). The context vector C_i for the output sequence Y_i is calculated by taking the weighted sum of the hidden states, h.

$$C_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j \tag{11}$$

Here T_x denotes the total number of hidden states and the weights α_{ij} are calculated by a softmax function as: $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x}\exp(e_{ik})}$ and $e_{ij} = a(s_{i-1}, h_j)$.

Where, e_{ij} is a fully connected neural network that measures how well the inputs around position j and outputs in position i match. The values of e_{ij} and α_{ij} are iteratively updated with every epoch of training.

In recent years, attention mechanisms have considerably improved RNNs' ability to handle long sequences and intricate dependencies, marking a significant advancement in the field (Dikshit et al., 2022). When integrated with the encoder-decoder-based LSTM model, the attention model produces remarkable results compared to the LSTM model alone (Ding et al., 2020). Although attention was initially developed for seq2seq time series problems, the concept is now applied to image-based tasks that involve object detection (Li et al., 2020), and semantic segmentation (Hu et al., 2021).

3.7. Transformer network

The Transformer network, proposed by Vaswani et al. (2017), is the most recent DL architecture that has gained attention after it surpassed the performance of the state-of-the-art LSTM models in time series forecasting/prediction problems. The accuracy of the RNN/LSTM model declines as the length of the input sequence increases since it becomes difficult for the model to capture the longer historical time lags with recurrence time steps, which is called the vanishing gradient problem. On the other hand, the transformer model discards the recurrence operations and utilizes a *self-attention* mechanism, a feature that could be leveraged for multi-step time-series prediction (Ahmed et al., 2022).

The transformer model comprises an encoder-decoder-based architecture (Fig. 11; Wu et al., 2020). The encoder block consists of an input layer, a positional encoding layer, and a series of encoders. The input layer is an FCNN that transforms the input sequence $(t_1, t_2, \cdots, t_k)$ into a d-dimensional vector. The positional encoding layer preserves the sequential information, a critical aspect when dealing with time series hydrological data.

Each encoder comprises two sub-layers – a self-attention and an FCNN sub-layer. The self-attention sub-layer measures the relevance score of each value at a particular time step with respect to all other values at the different time steps. These relevance scores capture interdependence relationships among the different time steps and are represented as attention vectors (Wu et al., 2020). The self-attention mechanism allows the network to focus on crucial historical lag time

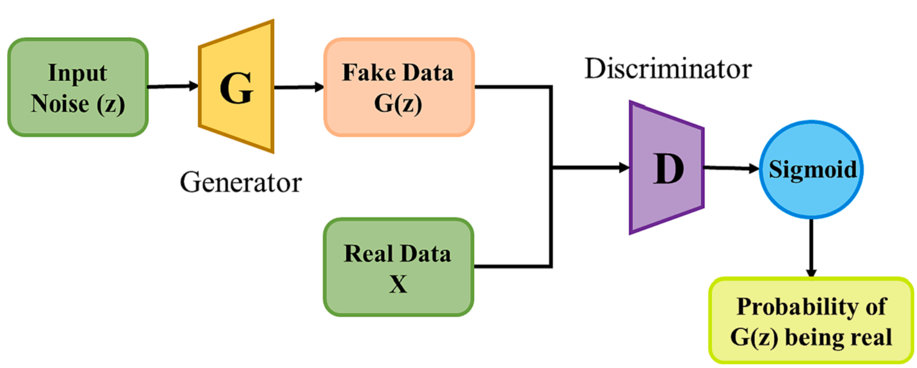

Fig. 8. Schematic diagram of Generative Adversarial Network (GAN) model.

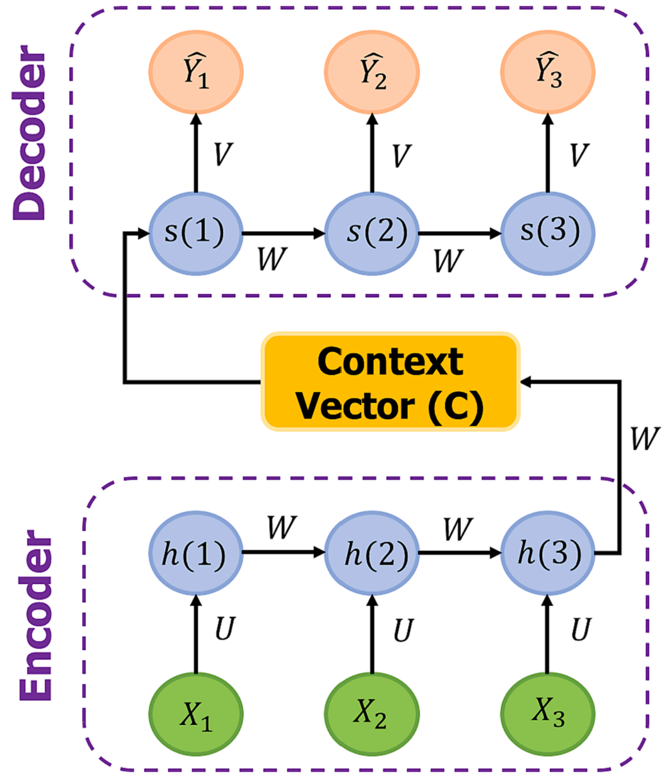

Fig. 9. The diagram depicts the encoder-decoder LSTM architecture designed for seq2seq prediction tasks.

steps for very long input sequences.

The decoder block comprises the input layer, the same number of decoder layers as the encoder block, followed by an output layer. In addition to the self-attention and FCNN sub-layers, each decoder layer consists of a third sub-layer called the encoder-decoder attention that applies the same self-attention operation over the last encoder output. A look-ahead masking scheme is employed in each decoder layer that efficiently helps train the transformer. It involves passing the decoder input sequences (t_k, t_{k+1}) to the decoder block by hiding the next sequence (t_{k+2}) that allows the network to learn to predict the next sequence correctly and iteratively update the weights using a loss function. This operation is carried out in each decoder and sent to the output layer that maps the output of the last decoder to the target time sequence using a linear activation function.

Further improvements of the Transformer model include the application of multi-head attention, which allows the model to capture different types of information from the input in parallel. In this way, the model can focus on different positions and create a more comprehensive representation of the input data (Vaswani et al., 2017), which could be beneficial in dealing with the multiscale and multivariate nature of hydrological processes. Although the applications of novel architectures (i.e., attention model and transformers) are very limited in hydrology, their demonstrated success in other disciplines (Braşoveanu and Andonie, 2020; Camps-Valls et al., 2020; Feng et al., 2021; He et al., 2022) suggests potential for exponential growth within the field in future.

3.8. Transfer learning (TL)

Transfer Learning (TL) offers an innovative solution to address the challenges of training DL models on limited datasets and reducing overall training time. This method leverages previously trained models by reusing their feature maps, or model parameters, to facilitate the learning process for new models (Weiss et al., 2016). TL assumes that feature maps, once learned, remain applicable across different tasks and

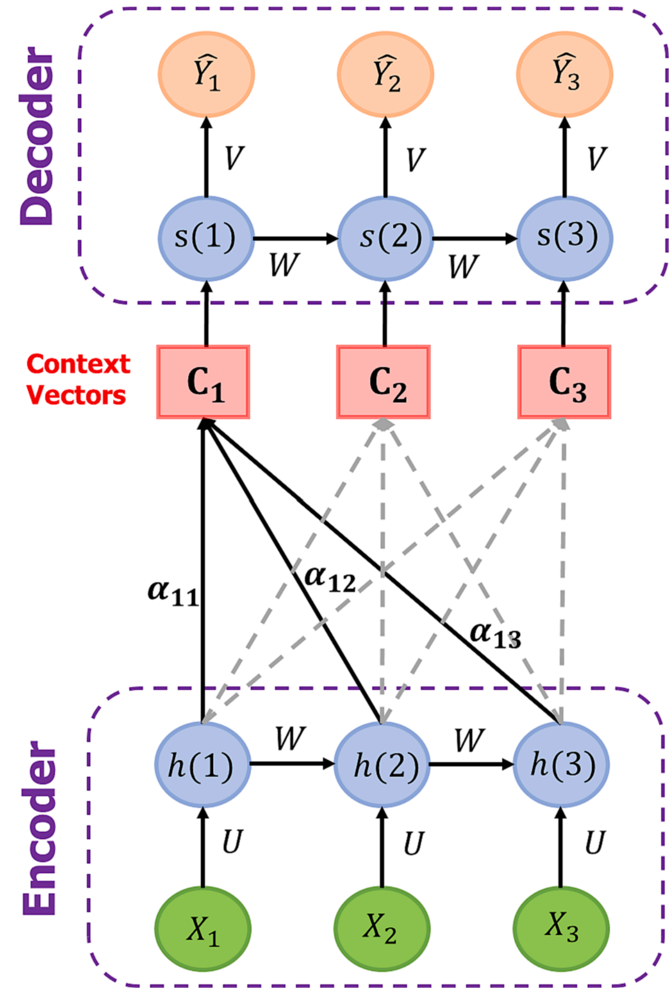

Fig. 10. The Attention model architecture based on an encoder-decoder model for seq2seq prediction problems.

datasets (Fig. 12). This ability to transpose learned features enhances the versatility and reusability of DL models, making them adaptable for analogous tasks on different datasets or for various tasks on a single dataset (Yu and Ma, 2021).

A classic application of the TL concept can be found in hydrological forecasting. For example, trained weights and biases from a model used for drought forecasting could serve as initial parameters for streamflow forecasting with a different dataset (Subramanian et al., 2022; Fig. 12). Furthermore, TL allows us to exploit feature maps extracted from a single dataset, like temperature data, for distinct tasks such as predicting and classifying heatwaves. This approach saves substantial training time by retaining the lower layers of the network while modifying the higher layers according to the task.

TL can also be especially beneficial in scenarios where training data is scarce. It leverages the knowledge gained from a DL model trained on a rich, labelled dataset and applies this knowledge to situations where training samples are limited. An excellent illustration of this is predicting runoff in ungauged catchments based on models trained in gauged areas with similar climate patterns, thereby broadening the model's practical applicability (Xu et al., 2023b).

4. Deep Learning: Future perspectives

4.1. Physics-Guided Deep learning (PGDL)

Hydrological models are quite complex, representing the physical

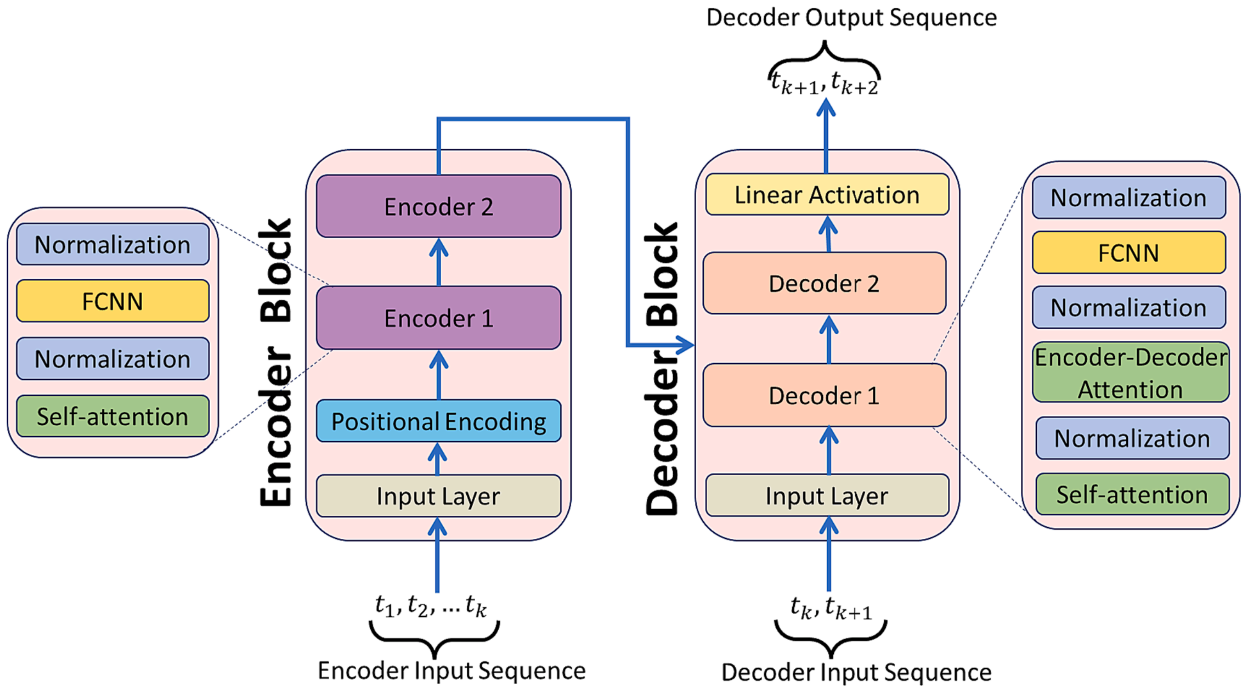

Fig. 11. The schematic representation of the encoder-decoder-based Transformer network for seq2seq timeseries prediction problems.

and conceptual relationships between multiple variables (processes) at different spatiotemporal scales. Traditional approaches to build hydrological models include physics-based concepts that aim to map the cause-effect relationship among the variables. These models use numerical simulations that encode physical laws as differential equations and numerical methods (Brunner and Simmons, 2012; Pinder and Gray, 2013; Yeh, 1986). However, these methods require significant computational resources as the complex dynamical processes require accurate discretization of the domain attributes and optimization of the constraints (Wang and Yu, 2022). Unfortunately, physics-based models often struggle to capture the complex and nonlinear hydroclimatic processes and nonstationary patterns that dynamically vary over space and time. On the other hand, data-driven DL models provide alternative tools for extracting useful information through complex pattern recognition and emulating nonlinear dynamics, thus limiting computationally demanding calculations (e.g., numerical approximation). Yet, one of the criticisms is that DL models may not capture the underlying laws of physics, which may lead to physically implausible and spurious predictions (Khandelwal et al., 2020; Shen et al., 2021).

Considering the above differences, it is essential to delve deeper into the comparison between DL/AI models and traditional physics-based models. DL models excel in capturing complex, nonlinear relationships in hydrological data. At the same time, physics-based models, despite their demand for high computational resources, incorporate the physical processes involved and are often guided by rigorous laws and principles. DL and physics-based models have shown considerable success in hydrology and water resources applications, albeit in different ways and under varying circumstances.

Recently, Physics-Guided Deep Learning (PGDL) or Theory-Guided Data Science (TGDS; (Ganguly et al., 2014; Karpatne et al., 2017) has gained attention by integrating the physical/processes based principles with deep neural networks. PGDL (supplementary Fig. S1) aims to maximize the benefits of physics-based and deep-learning models to handle scientific problems better. Combining physical mechanisms and data-driven approaches offers complementary strengths: data-driven methods extract valuable information from data, while physical models provide interpretability and generalization beyond the observation space. PGDL aims to enhance standalone deep learning models' physical consistency and generality, making it useful for the dynamical modeling of nonlinear and nonstationary processes and situations where system parameters change (Wang and Yu, 2022).

Karpatne et al. (2017) introduced a framework to combine domain knowledge with data-driven models for better physical consistency. The scientific principles can be incorporated into the design of data-driven approaches by selecting a suitable response function or redesigning the model architecture. This is very useful in hydrologic modeling as certain hydrological variables follow specific patterns or distributions or have a particular form of input–output relationship. Further, scientific knowledge can be used for identifying initial values or physics-guided regularization/optimization steps. Appropriate initial values can improve learning and generalization through pre-training networks and transfer learning. Few studies have modified the objective function to include scientific constraints like conservation of mass, energy, and momentum (Beucler et al., 2021; Daw et al., 2022). The DL model outputs can be post-processed using domain knowledge to improve physical consistency and interpretability.

Data-driven Deep Learning (DL) methods and physics-based models complement each other (Karpatne et al., 2017). DL methods excel in capturing the spatial and temporal dependencies and nonlinear patterns in data. At the same time, physics-based models provide a deeper understanding of the physical processes involved and offer interpretable results. The combination of these two models through hybrid modeling can lead to the acquisition of their individual strengths. There are two ways to construct a hybrid model. One approach is to use the outputs of the physical models as additional training inputs for the DL models, allowing the physical models to guide the learning process of the DL models with physically consistent training data. Another approach is to have the DL models predict specific intermediate steps in the physical models and feed the DL model outputs back into the physical models. This is especially useful when equations in the physical models are based on empirical relationships that can be difficult to interpret, such as Manning's formula. Modeling these intermediate relationships with DL can improve the prediction accuracy and correct the outputs of physics-based models. The DL model can learn the unknown parameterization of physics-based models, such as global hydrologic models (GHM), which will help improve their realizations (Yang et al., 2019b; Zaherpour et al.,

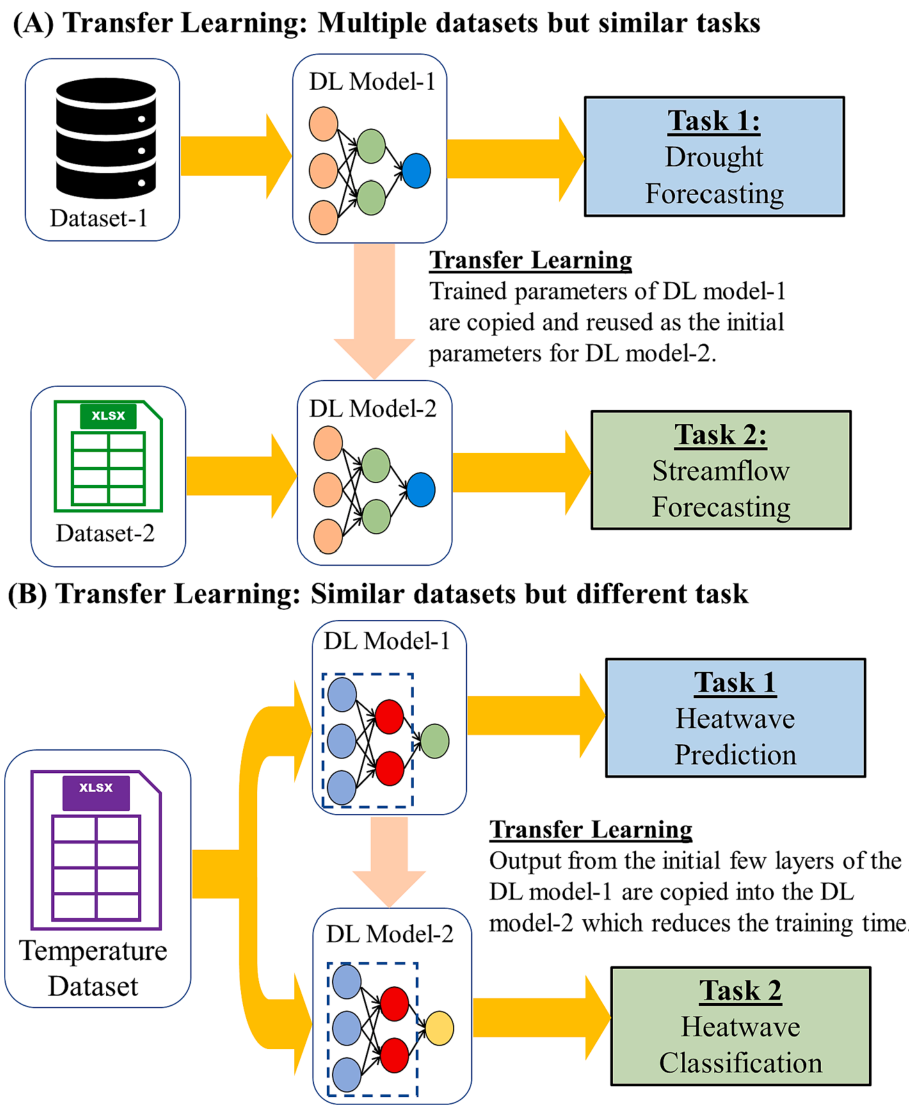

Fig. 12. An illustration of Transfer Learning in hydrology. (a) This panel demonstrates how transfer learning operates across similar tasks. It displays the process where weights and biases from a Deep Learning model trained on drought forecasting (Task-1) serve as initial parameters for streamflow forecasting (Task-2). This exchange of information between models enhances efficiency and saves computational resources. (b) Here, transfer learning across different tasks within the same dataset is depicted. Specifically, early layer feature maps from a model trained on heatwave prediction (Task-1) are reused by another model to classify heatwaves (Task-2). This demonstrates the utilization of shared information for cross-task learning, leading to improved model performance.

2019). Additionally, with a large number of physical parameters, it becomes expensive for the DL model to learn the optimal combination that maximizes the likelihood (Krizhevsky et al., 2012; Rasp et al., 2018).

In conclusion, while AI and physics-based models have their strengths and limitations, the synergy of these approaches, as evidenced by PGDL, suggests promising directions in advancing the hydrological modeling science.

4.2. Explainable Artificial Intelligence (XAI)

Although DL has gained success in numerous scientific fields, these models are sometimes criticized as “black boxes” as they don’t provide insights to understand how they make predictions (Castelvecchi, 2016). This is primarily due to their hierarchical nonlinear nature and over-parameterization — involving a large number of parameters and stacked with several layers. For any parametric statistical modeling, more parameters can capture the influence of diversely interacting variables. However, this shadows their interpretability, known as the accuracy-interpretability trade-off, where the model accuracy increases with more parameters but at the cost of model interpretability. Interpretability refers to the degree to which inference results of a DL network are predictable or understandable to humans. For example, the streamflow of a basin can be modeled as a linear function of several covariates in a naïve approach. In this case, the parameters of the multiple linear regression model are easily interpretable, and we can infer the relative contribution of covariates on the runoff generation. However, in reality, the runoff generation is influenced by a complex nonlinear relationship between hydrological variables under various physiographic conditions, requiring the use of nonlinear models with a large number of parameters (Konapala and Mishra, 2020; Woods and Sivapalan, 1999). This nonlinearity accounts for a more significant number of parameters, which hampers the model interpretability.

The complex network architectures of DL models involve many parameters set to gain higher accuracy; however, learning representations are difficult to extract and present in a human-readable form. Researchers have been continuously working to explain the behavior of the deep models in making decisions/predictions, a field called explainable AI (XAI; Ribeiro et al., 2016). Successful interpretation of deep models can help us to gain domain insights and extend our knowledge about unknown mechanisms, causation, and linkages. This motivates to go beyond DL as a knowledge discovery tool rather than a data-fitting model. Three main categories of XAI techniques are used to interpret the predictions of deep learning models:

1. Model-Agnostic Techniques: These techniques do not require internal knowledge of the model architecture and can be applied to any DL model. They aim to provide a global understanding of the model's behavior by focusing on the relationships between the inputs and outputs. Examples of these techniques include partial dependence plots (Friedman, 2001; Konapala and Mishra, 2020), individual conditional expectation plots (Goldstein et al., 2015), integrated gradients (Sundararajan et al., 2017), and saliency maps (Simonyan et al., 2014).
2. Model-Specific Techniques: These techniques are based on a specific model architecture and require an understanding of the model's internal workings. They offer a deeper understanding of the model's behavior and help identify the neurons and layers important for making the prediction. Examples of these techniques include layer-wise relevance propagation (Bach et al., 2015) and guided back-propagation (Springenberg et al., 2015).
3. Post-hoc Techniques: These techniques are applied after the model has been trained and provide an explanation for a specific prediction. They aim to highlight the specific input features that contribute the most to the model's prediction. Examples of these techniques include LIME (Local Interpretable Model-agnostic Explanations (Ribeiro et al., 2016) and SHAP (SHapley Additive exPlanations; (Lundberg and Lee, 2017).

Recent advances in DL research have also developed software tools for interpreting DL algorithms. TF-Explain is a TensorFlow-based interpretation algorithm focusing on features with gradient-based model agnostics techniques. Captum (Kokhlikyan et al., 2020) is based on PyTorch and works similar to TF-Explain.

5. Applications of DL in hydrology and water resources disciplines

During past decades, first-generation neural networks were widely used to solve a multitude of hydrological problems, such as water quality modeling (Singh et al., 2009), hydrological prediction/forecasting (Maqsood et al., 2004; Mishra et al., 2007; Mishra and Desai, 2006; Taylor and Buizza, 2002), remote sensing image object recognition/ classification (Mas and Flores, 2008; Tatem et al., 2001), water resources management (Iliadis and Maris, 2007; Jain et al., 2001; Kingston et al., 2005), and subsurface flow systems (Kerem Cigizoglu and Kisi, 2006; Lallahem et al., 2005; Shigidi and Garcia, 2003). Over recent years, the applications of DL in the hydrology discipline have emerged. The following sections provide an overview of recent applications of DL models, and a summary is provided in Table 2.

5.1. Time series modeling and forecasting

Hydrological time series forecasting involves extracting meaningful statistical information from sequential data for developing forecasting models. The forecasting models are widely used for various applications, such as predicting extreme hydroclimatic events, streamflow/runoff, and other hydrological variables (Coulibaly and Baldwin, 2005; Mujumdar and Kumar, 2012). The DL methods performed well compared to the statistical and stochastic models (e.g., Auto-Regressive Integrated Moving Average (ARIMA)) and processes-based hydrologic models. The commonly used traditional models have limitations, such as the choice of model formulation, parameter estimation, and assimilation scheme (Camps-Valls et al., 2021; Chawla et al., 2018). At present, the DL methods rapidly emerge as a legitimate alternative to classical time series modeling frameworks. An overview of DL applications for time series forecasting of floods, droughts, streamflow, soil-moisture, and weather forecasting are discussed.

5.1.1. Flood forecasting

Natural disasters like floods result in substantial loss of life, agricultural production, socioeconomic systems, and environmental sustainability (Bulti and Abebe, 2020; Mishra et al., 2022). Accurate flood modeling and forecasting are challenging in hydrology. Recent studies successfully applied DL methods and outperformed the process-based flood forecasting models in many cases. Wu et al. (2018) developed an LSTM model with an attention mechanism for hourly flood prediction, providing accurate and timely forecasts. Likewise, Fang et al. (2021) developed a hybrid model incorporating feature engineering with LSTM for flood susceptibility mapping. They implemented batch normalization and data augmentation, improving the model accuracy to approximately 94 %. In terms of streamflow prediction, Kratzert et al. (2019) made robust runoff predictions across 531 catchments throughout the continental United States using LSTM. Further, Feng et al. (2020) used a combination of Data Integration (DI) and LSTM, enhancing performance in basins with high autocorrelation. Ding et al. (2020) proposed an interpretable flood forecasting model that integrates a spatiotemporal attention mechanism with an LSTM network. The model outperformed all other benchmark models and the spatiotemporal attention layer, providing the importance of previous lagged values on flood prediction.

5.1.2. Drought forecasting

Drought prediction is paramount for water resources planning and management and for improving water security (Mishra and Singh, 2010). Dikshit et al. (2021) introduced a stacked LSTM model for drought predictions based on the Standardized Precipitation Evapotranspiration Index (SPEI) in New South Wales, Australia. In a subsequent study, Dikshit et al. (2022) improved the model by integrating an attention mechanism, yielding impressive accuracy for short- and long-term meteorological droughts. Xu et al. (2022) combined an autoregressive integrated moving average (ARIMA) model with LSTM for drought prediction based on SPEI over China. The hybrid model achieved better prediction ability than the selected ML and DL benchmark models. Mokhtar et al. (2021) applied two machine learning (Random Forest (RF), Extreme Gradient Boosting (XGBoost)) and two deep learning (CNN and LSTM) models for SPEI based meteorological drought prediction in Tibetan Plateau, China. The XGBoost and LSTM models showed the best performances in most of the scenarios.

5.1.3. Weather forecasting

The weather system is quite complex, and there is a high degree of uncertainty in predicting the state and conditions of the atmosphere in space and time. A hybrid spatiotemporal-LSTM model with self-attention scheme yielded promising results for weather forecasting, especially for moderate to heavy rainfall events (Zhang et al., 2022). Additionally, Wei and You (2022) implemented a combination of Discrete Wavelet Transform (DWT), LSTM, and Deep Convolutional Complementary Neural Network (DCCNN), improving forecasting accuracy for up to 4 months lead times. Chen et al. (2019) combined 3-D CNN and LSTM to develop a CNN-LSTM hybrid model that can capture the spatial correlation and temporal sequence of relations in typhoon progression. The proposed model showed superiority in predicting typhoon formation and intensity compared to the existing typhoon models. (Giffard-Roisin et al., 2020) proposed a DL model based on CNN integrating the past tropical cyclone trajectory information and atmospheric variable 3-D image fields such as wind and geopotential heights.

5.1.4. Soil moisture prediction

Statistical and ML models are commonly used for soil moisture prediction (Karthikeyan and Mishra, 2021; Martínez-Fernández and Ceballos, 2005; Yan et al., 2015). DL-based nowcasting of soil moisture has also produced remarkable results (Fang and Shen, 2020). Li et al.

Table 2
Application of DL models for various hydrological and water resources applications.

Applications	References	DL Model used	Significant Findings
Flood Forecasting	Hourly flood forecasting (Wu et al., 2018)	Context-aware LSTM with an attention mechanism	Better forecasting accuracy than MLP and LSTM models.
	Runoff prediction from short term extreme rainfall data (Li et al., 2021)	LSTM	Accurate prediction of runoff with low computational time.
	Flood forecasting (Ding et al., 2020)	LSTM with spatiotemporal attention mechanism	Outperformed selected process-based and ML models.
Weather Forecasting	(Zhang et al., 2022)	Spatiotemporal-LSTM with self-attention	Superior accuracy, especially for moderate to heavy rainfall events.
	(Zhang et al., 2022)	CNN	Improved spatial predictions, though accuracy was unchanged.
	(Chen et al., 2019)	3-D CNN + LSTM	Superior in predicting typhoon formation and intensity.
	(Giffard-Roisin et al., 2020)	CNN	Fast and accurate tracking of cyclone paths.
	(Ham et al., 2019)	CNN	Skillfully predicted the Nino 3.4 index for up to 18 months.
Streamflow prediction	(Saha et al., 2022)	CNN	Highest prediction accuracy for flood susceptibility maps.
	(Feng et al., 2020)	DI + LSTM	Improved prediction, especially in basins with high autocorrelation.
Soil moisture prediction	Short-term soil moisture forecasting (Li et al., 2022)	LSTM with an attention mechanism	Outperforms traditional ML and DL models; attention weights of the predictors and temporal dependencies provide valuable insights into the model's interpretation.
	Soil moisture modeling (Fang et al., 2017)	LSTM	DL-based dynamic modeling approaches can approximate soil moisture accurately with only two years of data, outperforming statistical methods.
	Multilayer soil moisture estimation (Karthikeyan and Mishra, 2021)	XGBoost trained region-wise and layer wise	The XGBoost algorithm effectively estimated soil moisture across various depths, capturing temporal dynamics and spatial variability. It revealed the relative importance of different factors in prediction and exhibited superior performance in deeper soil layers.
Remote-sensing applications	Object Detection and Classification (Yang et al., 2022)	CNN	Classified urban wetlands from high-resolution multispectral images with an accuracy of 90 %
	Landslide Detection (Tang et al., 2022)	Transformer-based DL model	Detected coseismic landslides from high-resolution remote sensing images and outperformed other models
	Land cover classification (Paul and Nagesh Kumar, 2018)	Mutual-information-based stacked autoencoder model	Used to extract spatial-spectral features from hyperspectral images for land cover classification
	LULC Classification (Zhang et al., 2019)	3-D CNN and 3-D DenseNet	Achieved over 99 % accuracy in LULC classification from hyperspectral images
	Hurricane detection (Kaur et al., 2022)	CNN	Used for hurricane damage assessment, achieving an accuracy of over 95 %
	Anomalous precipitation detection (Murakami et al., 2022)	Autoencoder model	Used to identify and classify anomalous precipitation events
Rainfall-runoff modeling	Runoff estimation (Xiang et al., 2020)	Encoder-decoder LSTM	Improved rainfall-runoff modeling accuracy.
	Runoff estimation (Jiang et al., 2020)	Hybrid Physics-RNN and 1D-CNN	Hybrid model outperformed individual LSTM and EXP-HYDRO.
Inverse Problem Modeling	Geostatistical inversion of geological media (Laloy et al., 2017)	Variational Autoencoder (VAE)	VAE outperformed traditional methods in a 2D steady flow case study.
	Geostatistical inversion of geological media (Laloy et al., 2018)	Spatial GAN	Spatial GAN required fewer training images and was faster than the VAE.
	Macro-dispersivity and hydraulic conductivity field mapping (Zhou et al., 2020)	CNN	Promising for mapping between complex subsurface structures and solute transport behavior.
Downscaling	Downscale surface soil moisture (Liu et al., 2022b)	LSTM	Efficiently addressed multi-scale and multi-source data challenges.
	Generating high-resolution daily precipitation data (Tu et al., 2021)	Hybrid WRF-CNN model	Downscaled 80 km resolution to 6 km in less time, with satisfactory results.
	Downscaling extreme temperature and precipitation (Wang et al., 2020b)	RNN-RandExtreme hybrid model	Outperformed ANN models in predicting extreme temperature and precipitation.
Water quality	Dissolved Oxygen level prediction (Zhi et al., 2021)	LSTM	Successfully captured DO level peaks and troughs during periods of low streamflow and DO data fluctuations.
	Short-term water quality prediction (Wan et al., 2022)	SOD-VGG-LSTM hybrid model	The hybrid model achieved high accuracy compared to other statistical and DL models in capturing extreme values.
	Predicting spatiotemporal variations of Dissolved Oxygen levels (Yu et al., 2020)	DL model	Data-driven DL model provided accurate predictions of DO level variations and hypoxic conditions.
	Water quality variables prediction (Bi et al., 2021)	LSTM-based encoder-decoder	Predicted water quality variables with satisfactory accuracy after denoising the data.
	Water quality prediction (Bi et al., 2023)	Hybrid Encoder-decoder based BiLSTM with an attention mechanism	The proposed model outperforms current state-of-the-art algorithms in prediction accuracy by efficiently handling noise, capturing long-term correlations, performing dimensionality reduction, and optimizing hyperparameters.
	River water quality (DO) prediction (Zhi et al., 2021)	LSTM	The LSTM model successfully predicts dissolved oxygen dynamics across minimally disturbed basins on a continental scale, leveraging sparse DO and daily hydrometeorology data.
Water level prediction	Surrogate Water level prediction in Yangtze River (Pan et al., 2020)	CNN-GRU model	Outperformed ARIMA, WANN, and LSTM models in predicting water levels.
	Daily water level variation prediction (Xu et al., 2023a)	Transformer model	Outperformed LSTM based on 7-day lead time predictions.

(continued on next page)

Table 2 (*continued*)

Applications	References	DL Model used	Significant Findings
	Surrogate groundwater level model (Cai et al., 2022)	Hybrid PGDL model	The model was more generalizable and robust than pure deep learning models.
	Groundwater patterns investigation (Clark, 2022)	SOM-LSTM model	The two-step modeling improved the predictive performance significantly.
Addressing multiscale problems	Multiscale Groundwater Level (GWL) Forecasting (Rahman et al., 2020)	XG Boost, RF, and Support Vector Regression, coupled with Wavelet Transforms (WT).	The coupling of ML approaches with WT improved the performance of GWL forecasting. XG Boost variants and RF provide an internal measure of variables' importance, making the models more interpretive over other black-box approaches such as SVR.
	Multiscale Soil Moisture Prediction (Liu et al., 2022a)	Multiscale LSTM model	The novel multiscale model achieved a record-breaking accuracy with a median correlation of 0.901 and RMSE of 0.034 m^3/m^3.

(2022) successfully combined the LSTM model with an attention mechanism to forecast soil moisture and temperature up to 7 days lead time. This innovative approach not only captures the attention of the predictors, but also incorporates self-attention to account for temporal dependencies. The results demonstrated that this model outperforms a majority of traditional machine learning and deep learning models. Furthermore, the attention weights of the predictors and temporal dependencies provide valuable insights into the model's interpretation, aligning with existing physical knowledge of soil moisture and temperature forecasts. Ahmed et al. (2021) implemented a hybrid DL architecture (CEEMDAN + CNN + GRU) to predict the remotely sensed surface soil moisture (SSM) for up to 30 days lead time. The results demonstrate that the proposed model can successfully forecast surface soil moisture compared to benchmark models.

5.2. Remote sensing image-based applications

Remote sensing allows for monitoring hydrological variables and processes across time and space. The current remote sensing-based applications in hydrology can be broadly classified into three categories: (1) object detection and classification, (2) land use land cover classification, and (3) change and anomaly detection, which are discussed in the following section.

Object detection and classification in remote sensing images have undergone significant advancements with the advent of DL techniques. This process, which entails differentiating and categorizing objects within these images, was previously fraught with challenges. A crucial issue was the difficulty of identifying rotation-invariant features in raw image data before the application of data augmentation. However, introducing a rotation-invariant layer to the CNN can address this limitation, which improves multiclass object detection (Cheng et al., 2016). Similarly, an object-based CNN model integrated with an autoencoder mechanism can enhance high-level feature extraction (Jiang, 2018). The CNN model achieved an impressive accuracy rate over 95 % using data augmentation and fractal net evolution.

DL architectures, notably autoencoders and CNN are proficient in extracting high-level spatial features from remote-sensing images, enhancing the LULC classification process (Maggiori et al., 2017). For example, Xing et al. (2018) utilized a deep CNN (DCNN) with a pre-trained VGG-16 network for LULC classification using geo-tagged photos, marking a notable improvement in classification accuracy compared to traditional methods. Similarly, DL models have been utilized in change and anomaly detection to identify variations and abnormalities from remote sensing data, such as cloud detection or hurricane damage assessment. Jeppesen et al. (2019) developed a Remote Sensing Network (RS-Net), a DL model, especially for detecting clouds from remote sensing images. RS-Net uses spatial and spectral signatures and is trained/validated with Landsat-8 Biome and Spatial Procedures for Automated Removal of Cloud (SPARCS) data. The RS-net model delivers the highest accuracy even over snowy and icy regions. Additionally, the model performs well for smaller satellite images with limited multispectral capabilities.

5.3. Rainfall-Runoff modeling

Rainfall-runoff modeling is among the most important steps in hydrologic modeling and flood prediction. Traditional rainfall-runoff modeling approaches commonly involve statistical methods and physics-based hydrological models (Blöschl, 2006; Coles et al., 2003). However, the challenges remain while dealing with non-linearity and complex temporal dependencies inherent in hydrological processes. DL models can capture complex nonlinear relationships and temporal dependencies; thus, they have emerged as a promising approach to enhance rainfall-runoff modeling accuracy.

Kratzert et al. (2018) employed an LSTM model to generate streamflow from 241 catchments in the CAMELS database. They used various meteorological forcings data and observed discharge data as the input to the LSTM model. The models trained with a collective group of basins demonstrated superior performance, especially in snow and precipitation-dominated regions. Han and Morrison (2022) proposed using DL models as a post-processor to correct the outputs of traditional hydrologic models. They developed an encoder-decoder LSTM model to enhance the National Water Model's predictive performance for hourly runoff forecasts. The results showcased a dramatic reduction in root mean squared error, highlighting the potential of DL in enhancing runoff estimates.

5.4. Modeling Inverse problems

Inverse problem modeling is a significant area within groundwater studies, often approached as a high-dimensional inversion problem. Traditional approaches like Markov Chain Monte Carlo and Ensemble Kalman Filtering can be resource-intensive and suffer from the curse of dimensionality (Evensen, 1994; Vrugt, 2016). DL can efficiently handle high-dimensional problems and capture the critical spatiotemporal features inherent in inverse problem modeling.

Mo et al. (2019) applied a deep autoregressive neural network using a CNN-based encoder-decoder framework to create surrogate models for contaminant transport problems. The model was capable of handling high dimensionality with improved fidelity. Despite this, the study suggested that the accuracy and generalizability of the network could be enhanced by using more training samples. In another interesting application, Wu et al. (2019) used an image-based framework employing CNN to model the effective diffusivity of 2D porous media. The results demonstrated improved accuracy with less computational cost than traditional Lattice Boltzmann simulations.

5.5. Uncertainty quantification

DL has emerged as a compelling tool in quantifying uncertainty and can significantly enhance the accuracy and reliability of predictions. For instance, Klotz et al. (2022) explored using DL to estimate uncertainties in hydrological predictions, specifically rainfall-runoff modeling. The results highlighted improved mixture density networks (MDN) performance in terms of reliability and accuracy in uncertainty estimates than the Monte Carlo dropout method, especially for low- and high-flow scenarios. The study also underscores the potential for further advancements in model development, metrics, and benchmarking for uncertainty estimation, urging community-wide collaboration to overcome these challenges. Abbaszadeh Shahri et al. (2022) introduced a new approach, Automated Random Deactivating Connective Weights (ARDCW), for estimating uncertainty in deep learning models for groundwater table (GWT) predictions. This technique employs randomly turned-off weights, enhancing model predictability without altering optimization processes. Notably, ARDCW outperformed the traditional models in performance, demonstrating its efficacy in a real-world project in Stockholm, Sweden.

5.6. Downscaling

Downscaling large-scale hydro-meteorological variables to a local scale is crucial for regional impact assessment (Tripathi et al., 2006). DL, especially CNN, has become an advantageous tool due to their ability to capture the spatial features of spatiotemporal datasets. Pan et al. (2019) enhanced precipitation downscaling by applying a stacked CNN to extract critical circulation features. They used predictors like geopotential height and precipitable water at a 3-hour frequency, outperforming other weather models and enhancing precipitation-related information retrieval. On a different approach, Wang et al. (2021) leveraged a deep CNN-based Super Resolution Deep Residual Network (SRDRN) for downscaling daily precipitation and temperature. This network excelled in capturing spatiotemporal patterns and statistical characteristics, thereby accurately reconstructing temperature and precipitation extremes across different locations. In a recent study, Gavahi et al. (2023) applied DL-based convolutional neural network architecture to merge and downscale multiple user-defined precipitation products using targeted rain gauge observations.

5.7. Water quality modeling and monitoring

Reliable water quality prediction can minimize water-borne diseases (Mishra et al., 2021) and improve environmental flow monitoring, water security, and the sustainability of the stream ecosystem ((Alnahit et al., 2022)). Different methods are employed to predict depth-wise lake and river water temperatures, dissolved oxygen (DO) levels, and other non-point pollutant sources. In a lake water temperature modeling study, Daw et al. (2022) utilized a physics-based neural network model for lake water temperature prediction, providing improved results over process-based models for two major U.S. lakes. Similarly, Willard et al. (2021) applied *meta*-transfer learning approaches to predict depth-specific lake water temperatures for unmonitored lakes, leveraging models from well-monitored lakes. Their findings suggest that the PGDL model outperforms the standard model, particularly in regions with sparse monitoring.

Moreover, accurate prediction of DO levels is vital for aquatic organisms, which is often challenging due to the sparse availability of DO data. Zhi et al. (2021) developed an LSTM model to forecast DO levels across numerous undisturbed catchments, showing promising results, particularly during periods of low streamflow. On the other hand, Barzegar et al. (2020) designed a hybrid CNN-LSTM model for predicting DO and chlorophyll-a levels, surpassing standalone models and traditional ML approaches. This hybrid model was particularly effective in capturing DO level fluctuations. Willard et al. (2021) adopted *meta*-transfer learning approaches for correctly predicting the depth-specific lake water temperature in unmonitored lakes by reusing the models from well-monitored lakes across the Western United States. They applied *meta*-transfer learning to the process-based general lake model (MTL-PB) and physics-guided deep-learning model (MTL-PGDL). The results suggested that the MTL-PGDL model performed better than the MTL-PB model. Even for regions with sparsely monitored lakes, the MTL-PGDL model outperformed the standard PGDL model.

5.8. Surface and ground water level prediction

Accurate forecasting of surface and groundwater levels is crucial for effective resource management, environmental monitoring, agricultural planning, and drought management (Liu et al., 2021). Recently, hybrid DL models blended between different DL architectures and physics-based models have been found to perform well compared to traditional methods. For example, Xu et al. (2023) demonstrated the effectiveness of a transformer model in simulating daily water level variations in Poyang Lake, surpassing the LSTM model's up-to-7-day lead time predictions. Similarly, Barzegar et al. (2021) found that a CNN-LSTM hybrid model outperformed traditional machine learning models in predicting water levels in North American lakes. In groundwater level prediction, Cai et al. (2022) proposed a hybrid PGDL model that outperformed pure DL models in accuracy and generalizability for measuring groundwater level fluctuations. Clark (2022) employed a combination of Self-Organizing Maps and LSTM models to classify groundwater level time series into unique temporal patterns, leading to significantly improved predictive performance.

5.9. Applications of Physics-Guided Deep learning (PGDL)

In the past few years, there has been increasing usage of PGDL in hydrology and similar fields such as geoscience and climate science, addressing several key challenges related to image classification (Huang et al., 2021), hydrological system modeling (Jia et al., 2021; Wang et al., 2020a; Xie et al., 2021), time series forecasting (Deman et al., 2022; Liu et al., 2022b), and anomaly/change detection (Zhong et al., 2019). Applications of DL in subsurface flow through porous media are limited due to large-scale heterogeneity, and obtaining adequate amounts of data is often challenging due to the difficulty in installing and maintaining the sensors. In this scenario, domain knowledge can complement the data-driven models to develop hybrid DL models. For example, Tartakovsky et al. (2020) developed a PGDL model for estimating hydraulic conductivity and learning its constitutive relationships with capillary pressure in subsurface flow. They tested the model for estimating the unknown space-dependent diffusion coefficients from a linear diffusion equation for saturated flow in a heterogeneous medium. They also evaluated the constitutive relationship in a nonlinear diffusion equation for unsaturated flow in a homogeneous medium. They trained a DNN model that used physical knowledge from the partial differential equations and data from ground-based observations. The proposed hybrid model showed superior performance over the physics-based and the standalone DNN models in evaluating the hydraulic conductivity by as much as 50 %. He et al. (2021) employed a theory-guided CNN architecture to model the contaminant transport in subsurface flow. The addition of physical constraints to DL complemented the predictive power and generalization of the CNN model alone. The results demonstrate that the PGDL approach can capture the localized features, is robust, and is less time-consuming.

Read et al. (2019) used the PGDL modeling framework to improve the predictions of lake water temperature based on LSTM. They modified the LSTM model by adding a penalty term to the LSTM loss function for violating the conservation of energy. The PGDL predictions showed a 0.5 °C lower root mean square error (RMSE) relative to a physics-based model alone. The proposed model showed higher generalization capability and scalability (the model can be used for predictions in many other lakes without compromising accuracy). Flood risk assessment using runoff data from global hydrologic models (GHM) is a popular concept in hydrology but is often inaccurate and needs further improvements. Yang et al. (2019) evaluated flood simulations from the CaMa-flood model based on the inputs from GHMs and compared their performance with the LSTM. They designed the LSTM network to receive meteorological forcing inputs from the CaMa-flood model to output daily streamflow. The proposed PGDL model demonstrated excellent performance over the CaMa-flood model, especially capturing the amplitudes of peak flood discharge. The authors urged that integrating physics-based models with DL can be a powerful tool for more robust and confident flood risk assessments. Xie et al. (2021) adopted PGDL to investigate the effects of extreme events and monotonic relationships in the simulation of rainfall-runoff processes across CONUS. They accommodate the effects of losses from the hydrological processes in the optimization of the objective function of the LSTM DL model. In addition, the synthetic outputs of the physical mechanisms were further passed to LSTM, which improved the simulation of the flood peaks during heavy storm events. The proposed approach achieved greater physically consistent estimates (avoiding negative values and capturing the flood peaks).

5.10. *Applications related to explainable Artificial Intelligence (XAI)*

In recent years, hydrologists have started using various interpretation algorithms to interpret the black-box nature of the DL models. Jiang et al. (2022) investigated different flooding mechanisms from an LSTM model trained on meteorological forcings, such as temperature, precipitation, and daily streamflow across CONUS. They used a model agnostic interpretation method (expected gradient; Erion et al., 2021) to reveal three dominant flooding patterns: snowmelt induced, recent rainfall induced flooding, and historical rainfall induced flooding. They used another model-specific interpretation method (additive decomposition; Du et al., 2019) to understand the internal hidden layers' output from various gates of the LSTM model. Althoff et al. (2021) combined ML and conceptual hydrologic models to create hybrid models to improve the streamflow predictions across three gauging stations within the Brazilian Cerrado biome. They used XAI techniques to reveal how the data-driven component of the hybrid model handles the runoff routing. They demonstrated the superior performance of the hybrid model over the conceptual hydrologic model and XAI untangled soil-moisture as the dominant factor for predicting streamflow, which agrees with the previous studies. Dikshit and Pradhan (2021) investigated the relative contribution of various climate indices and meteorological variables in drought prediction across different drought conditions and drought events using various Shapely additive explanations (SHAP; Lundberg and Lee, 2017) models. They reported that their LSTM model achieved good accuracy, and the SHAP models could correctly interpret the various drought mechanisms. Wang et al. (2022) used several ML and DL models to predict the NH4 + -N concentration in the Xiaoqing estuary, China. They used shapely additive explanation methods to interpret the DL model outputs and to understand the role of upstream of the river in the estuary. The XAI methods revealed that two stations monitoring water quality in the upper reaches of the river are mainly responsible for the water quality in the estuary. The authors asserted that the SHAP methods are conducive to understanding the direction and magnitude of the influence of input covariates on the estuary water quality.

6. Challenges and opportunities of DL applications in hydrology and water resources

DL has led to unprecedented success in computer vision in natural language processing (NLP), image classification, speech recognition, and language translation. However, implementing DL methods in water resources disciplines has been an emerging topic over the past few years. This section highlights challenges and opportunities for DL applications in water resources disciplines.

- Multi-source, Multiscale, and High-dimensional Data

Data sets used in hydrology and water resources are collected from various sources (e.g., observed, reanalysis, climate models, sensors), often inconsistent in spatiotemporal resolutions. Hydrologic modeling and robust quantification of spatiotemporal extreme events (e.g., droughts and floods) require a high-dimensional learning environment to capture the nonlinear interaction between multiple climate and catchment processes that evolve at different spatial and temporal scales (Konapala and Mishra, 2020). Multi-source and multiscale high-dimensional input data can make the learning task (e.g., computational and statistical learning) more difficult in a complex learning environment. DL models have the potential to advance water resources-related research by capturing the interaction between different hydro-climatic processes within high-dimensional learning and complex environments.

- Poor Data Quality and Lack of Labeled Dataset

Data mining and acquisition is a foremost challenge in applying DL/ML models to perform a hydrological modeling task. In computer vision exercises such as Natural Language Processing, the availability of a large volume of the labeled dataset makes the supervised learning task easier. However, hydrology-related research often suffers from a lack of labeled data sets. With little labeled data, a supervised learning task may lead to erroneous results as it may not decode enough feature maps (patterns) from the data. Besides the quantity of labeled data, the quality is much more crucial (Kusiak, 2017), as the poor quality of labeled data incurs noisy datasets and missing values, which can substantially affect the accuracy of the model. Therefore, an important practice is to maintain both quality and quantity of labeled dataset before performing a hydrological task. Additionally, in situations with limited labeled data or poor data quality, it is recommended to use unsupervised models (e.g., autoencoders).

- Model Interpretability

The black-box perception of the DL models is one of the major criticisms. Which means generating the final outputs (target) without a proper understanding of processes, interactions, and feedback between variables associated with the outputs. The black-box models are challenging to interpret, and the degree of interpretability depends on the level of model complexities. Therefore, the way forward is to design inherently interpretable DL models by appropriately explaining the features involved in model development (Rudin, 2019). For example, interpreting the DL model outputs with respect to the weights associated with neurons, identifying the set of important weights, and quantifying the role of inputs based on their corresponding weights. The DL models also do not explain the cause-effect relationships among hydrological variables in hydrological processes. One of the key limitations of decoding the connection between inputs and weights is primarily associated with the compressed information carried by the neurons after applying nonlinear activation functions. DL model interpretability is likely to be an important area of research in the future, and appropriate metrics should be developed to optimize the models based on high accuracy versus high interpretability or to optimize both of them.

- Integrating Data-driven DL models with Physics-based models

Although DL models have demonstrated significantly high accuracy in recent years, they often lack interpretability and produce physically inconsistent estimates. DL models are known for deriving data-definitive solutions for a task without truly understanding the underlying physical mechanisms, cause-effect relationships, and interconnected systems/processes. Reichstein et al. (2019) proposed using PGDL approaches that integrate domain knowledge from physics-based models with DL models to produce physically consistent predictions to address this issue. PGDL approaches have gained considerable attention in hydrology and water resources applications due to their ability to improve accuracy, better generalization, and work well with limited training data.

- Addressing Nonstationarity and Uncertainties

Hydrological processes evolve with space and time; in other words, they are often not stationary. Nonstationarity refers to changes in statistical measures (e.g., mean and variance) of hydrological variables/phenomena over time (Chandra et al., 2015; Cheng et al., 2014; Rubin et al., 1995). The state-of-the-art DL models may not capture the effects of nonstationarity since they presume the training and testing data distributions are identical. Modeling nonstationary phenomena requires novel DL architectures dealing with statistically time-varying parameters. One approach to model them can be the PGDL approach, where the time-varying trend components can constrain neural network training.

Modeling uncertainty is an inherent challenge in hydrologic phenomena or processes. These uncertainties arise either due to the parametrization of the hydrological model or related to the data (e.g., due to the use of land surface models or general circulation models). While advanced DL architectures such as Bayesian Deep Learning have been applied to model uncertainties in other disciplines (Abdar et al., 2021; Kendall and Gal, 2017; Wang and Kadıoğlu, 2021), there is a scope to implement such concepts for hydrology and water resources disciplines.

- *Internet of Things (IoT)* and *Web-based Data Analytics Framework for Real-time Applications*

One of the significant challenges in hydrology and water resources management is the need for real-time monitoring and decision-making. Traditional data analysis and modeling methods may not be able to keep up with the high volume and velocity of data generated in real-time, making it difficult to extract insights and make informed decisions quickly. With the advancement in graphical processing units (GPU), there is ample scope for using web systems for large-scale data processing, analysis, and visualization for real-time hydrological applications. The pre-trained benchmark networks can be created using massive datasets and tested based on various stakeholder applications in a transfer learning framework.

The Internet of Things (IoT) is an emerging technology that can be leveraged in hydrological and water resources applications. With IoT, it is possible to deploy a large number of sensors to capture real-time data and transmit it to a central location for analysis and decision-making (Abdul Ghapar et al., 2018). By combining the power of IoT with web-based data analytics frameworks, it is possible to build systems that can handle the large volume of data generated by hydrological and water resources systems. The IOT-enabled web-based framework also enables collaboration and sharing of data and models, improving the decision-making process's accuracy and efficiency.

- High Computational Demand

DL models typically require a lot of computational resources, including high-performance computing (HPC) systems and graphics processing units (GPUs) for training and decision-making. As hydrological datasets become huge and more complex, the computational demands of DL models also increase. For example, analyzing the global scale data at a fine resolution of 10 km can be a significant challenge in terms of computational (time) demand as the number of grids will be approximately 3600x1800 pixels. Supercomputing and advanced hardware such as GPUs and TPUs can handle a massive number of operations per cycle. Still, they may not provide the level of computational power required for fine-resolution water resource applications (Reichstein et al., 2019). However, recent advancements in cloud computing and parallel processing techniques have made it easier to train and deploy DL models on a larger scale.

Additionally, researchers have been developing more efficient DL architectures and optimization algorithms to reduce the computational demand while maintaining high accuracy (Thompson et al., 2022). The use of transfer learning techniques can significantly reduce the training time and computational cost without sacrificing performance. As DL models continue to improve and evolve, the computational demand will likely become less of a barrier to their adoption in hydrology and water resources research.

- *Model Generalization and Transferability*

DL models often face the challenge of transferability and generalization across different regions and climatic conditions. DL models trained on data from one region may not perform well in another region due to differences in hydrological processes, data characteristics, soil properties, vegetation, and human influence. Consequently, developing robust transfer learning and domain adaptation methods could significantly enhance the ability of models to generalize across regions (Pan and Yang, 2010). Transfer learning could allow models to discern patterns from various climatic areas, potentially improving performance in unseen or underrepresented regions. However, implementing these techniques requires careful consideration of differences in feature space, distribution, and even structural differences in hydrological processes across regions. Therefore, significant research opportunities exist in developing effective transfer learning methods for hydrological modeling, universally applicable across diverse regions and conditions.

- *Data Augmentation and Synthetic Data Generation*

Methods such as data augmentation and synthetic data generation, which have proven useful in other disciplines, are notably challenging to implement in hydrology. For instance, creating new climate scenarios based on existing simulations is a common practice, but its validity for training DL models remains unclear. The use of DL models to generate realistic synthetic data for augmenting training datasets could potentially enhance model performance (Shorten and Khoshgoftaar, 2019). However, creating synthetic data that accurately represents complex hydrological processes raises many questions about the synthetic data's quality, physical validity, and potential impact on model training. Additionally, it poses questions about the computational resources and expertise needed for such data generation, potentially limiting its implementation. Therefore, the development of effective and valid techniques for data augmentation and synthetic data generation in hydrology presents both a significant challenge and an opportunity for improving DL model performance.

7. Conclusions

DL models have the potential to handle and discover hidden patterns in complex and high-dimensional data sets, which are valuable for hydrology and water resources disciplines. In this review, we highlighted the recent developments in DL concepts, methodologies, and applications. The following conclusions can be drawn from this study:

(a) DL methods have gained momentum for predicting various hydrologic fluxes in the water cycle, hydroclimatic extreme events, and extracting meaningful information for various water resource applications. The state-of-the-art RNN architectures like LSTM, variants of CNNs, and hybrid models based on the combination of these models can provide superior configurations that can outperform the traditional conceptual and statistical models.
(b) DL methods can complement traditional physical (processes) based models; for example, the Physics-Guided Deep Learning (PGDL) modeling framework combines domain and processed-based knowledge and DL concepts, a valuable tool for various applications. Incorporating domain (expert) knowledge related to the physical understanding of hydrological processes can reduce the search space of model parameters, leading to more accurate simulations, physically consistent predictions, and improved generalization (Karpatne et al., 2017; Wang and Yu, 2022). By incorporating physical constraints, PGDL offers physically

consistent and robust solutions that outperform state-of-the-art DL models in many applications.

(c) It is essential to comprehend and retrace how the DL algorithm came to a result or a decision. Therefore, it is crucial to understand the DL models' architecture better instead of trusting them unquestioningly (i.e., referred to as a "black box"). The concept of XAI allows the users to comprehend and gain confidence in the results and output created by DL algorithms. XAI can be used as a scientific knowledge discovery tool rather than just a data-fitting tool. Using techniques like saliency maps, activation visualization, and attribution methods, XAI aims to make DL models more transparent and interpretable in dealing with complex hydrological processes evolving over different landscapes.

(d) The applications of the Attention Model and the Transformer Network models in hydrology are currently limited. These concepts demonstrate significant advantages over traditional RNN/LSTM and CNN models regarding computational efficiency and accuracy. With the rapid growth in related fields and the availability of advanced computing tools like GPU/TPU, we expect that hydrologists will increasingly adopt these innovative models to tackle time-series related problems.

(e) DL methods can help deal with hydro-climatic non-stationary time series that often witness 'data bursts,' seasonality, structural break, and heteroscedasticity, a key challenge in traditional statistical predictions. Therefore, new concepts and models must be developed to adapt to the abrupt (unforeseen) changes in hydroclimatic time series for decision-making. DL models offer promising opportunities by integrating domain knowledge and hidden patterns in underlying data sets.

(f) Real-time decision-making is critical in the context of climate extremes for minimizing their risk in water resources related sectors. Integrating DL with emerging technologies such as the Internet of Things (IoT) can provide real-time data analytics and decision-making capabilities for various applications by water resources communities. DL models can be trained and deployed in real-time web-based frameworks for applications such as flood forecasting and water resource management using IoT devices to collect and transmit data from various sources such as sensors and satellites.

CRediT authorship contribution statement

Kumar Puran Tripathy: Conceptualization, Visualization, Writing – original draft. **Ashok K. Mishra:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This study is supported by the National Science Foundation (NSF) award # 1841629 and 1653841.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2023.130458.

References

M. Abadi A. Agarwal P. Barham E. Brevdo Z. Chen C. Citro G.S. Corrado A. Davis J. Dean M. Devin S. Ghemawat I. Goodfellow A. Harp G. Irving M. Isard Y. Jia R. Jozefowicz L. Kaiser M. Kudlur J. Levenberg D. Mane R. Monga S. Moore D. Murray C. Olah M. Schuster J. Shlens B. Steiner I. Sutskever K. Talwar P. Tucker V. Vanhoucke V. Vasudevan F. Viegas O. Vinyals P. Warden M. Wattenberg M. Wicke Y. Yu X. Zheng TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems 2016 https://doi.org/10.48550/arXiv.1603.04467.

Abbaszadeh Shahri, A., Shan, C., Larsson, S., 2022. A novel approach to uncertainty quantification in groundwater table modeling by automated predictive deep learning. Nat. Resour. Res. 31, 1351–1373. https://doi.org/10.1007/s11053-022-10051-w.

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V., Nahavandi, S., 2021. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Inf. Fusion 76, 243–297. https://doi.org/10.1016/j.inffus.2021.05.008.

Abdul Ghapar, A., Yussof, S., Bakar, A., 2018. Internet of Things (IoT) Architecture for Flood Data Management. Int. J. Future Gener. Commun. Netw. 11, 55–62. https://doi.org/10.14257/ijfgcn.2018.11.1.06.

Abrahart, R.J., See, L.M., 2007. Neural network modelling of non-linear hydrological relationships. Hydrol. Earth Syst. Sci. 11, 1563–1579. https://doi.org/10.5194/hess-11-1563-2007.

Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol. 407, 28–40. https://doi.org/10.1016/j.jhydrol.2011.06.013.

Ahmed, S., Nielsen, I.E., Tripathi, A., Siddiqui, S., Rasool, G., Ramachandran, R.P., 2022. Transformers in Time-series Analysis: A Tutorial.

Allamano, P., Croci, A., Laio, F., 2015. Toward the camera rain gauge. Water Resour. Res. 51, 1744–1757. https://doi.org/10.1002/2014WR016298.

Alnahit, A.O., Mishra, A.K., Khan, A.A., 2022. Stream water quality prediction using boosted regression tree and random forest models. Stoch. Environ. Res. Risk Assess. 36, 2661–2680. https://doi.org/10.1007/s00477-021-02152-4.

Althoff, D., Bazame, H.C., Nascimento, J.G., 2021. Untangling hybrid hydrological models with explainable artificial intelligence. H2Open J. 4, 13–28. https://doi.org/10.2166/h2oj.2021.066.

Amaranto, A., Munoz-Arriola, F., Solomatine, D.P., Corzo, G., 2019. A Spatially Enhanced Data-Driven Multimodel to Improve Semiseasonal Groundwater Forecasts in the High Plains Aquifer, USA. Water Resour. Res. 55, 5941–5961. https://doi.org/10.1029/2018WR024301.

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein GAN. https://doi.org/10.48550/arXiv.1701.07875.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large Area Hydrologic Modeling and Assessment Part I: Model Development1. JAWRA J. Am. Water Resour. Assoc. 34, 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x.

Asadi, A., Safabakhsh, R., 2020. The Encoder-Decoder Framework and Its Applications, in: Pedrycz, W., Chen, S.-M. (Eds.), Deep Learning: Concepts and Architectures, Studies in Computational Intelligence. Springer International Publishing, Cham, pp. 133–167. https://doi.org/10.1007/978-3-030-31756-0_5.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., Samek, W., 2015. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. PLOS ONE 10, e0130140.

D. Bahdanau K. Cho Y. Bengio Neural Machine Translation by Jointly Learning to Align and Translate 2016 https://doi.org/10.48550/arXiv.1409.0473.

Baldi, P., Sadowski, P.J., 2013. Understanding Dropout. Advances in Neural Information Processing Systems. Curran Associates Inc.

Ballard, D.H., 1987. Modular learning in neural networks, in: Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1, AAAI'87. AAAI Press, Seattle, Washington, pp. 279–284.

Barzegar, R., Aalami, M.T., Adamowski, J., 2020. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. Stoch. Environ. Res. Risk Assess. 34, 415–433. https://doi.org/10.1007/s00477-020-01776-2.

Barzegar, R., Aalami, M.T., Adamowski, J., 2021. Coupling a hybrid CNN-LSTM deep learning model with a Boundary Corrected Maximal Overlap Discrete Wavelet Transform for multiscale Lake water level forecasting. J. Hydrol. 598, 126196 https://doi.org/10.1016/j.jhydrol.2021.126196.

Best, M.J., Pryor, M., Clark, D.B., Rooney, G.G., Essery, R.L.H., Ménard, C.B., Edwards, J.M., Hendry, M.A., Porson, A., Gedney, N., Mercado, L.M., Sitch, S., Blyth, E., Boucher, O., Cox, P.M., Grimmond, C.S.B., Harding, R.J., 2011. The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. Geosci. Model Dev. 4, 677–699. https://doi.org/10.5194/gmd-4-677-2011.

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., Gentine, P., 2021. Enforcing Analytic Constraints in Neural-Networks Emulating Physical Systems. Phys. Rev. Lett. 126, 098302 https://doi.org/10.1103/PhysRevLett.126.098302.

Beven, K., 1989. Changing ideas in hydrology — The case of physically-based models. J. Hydrol. 105, 157–172. https://doi.org/10.1016/0022-1694(89)90101-7.

Bhatta, B., Shrestha, S., Shrestha, P.K., Talchabhadel, R., 2019. Evaluation and application of a SWAT model to assess the climate change impact on the hydrology of the Himalayan River Basin. CATENA 181, 104082. https://doi.org/10.1016/j.catena.2019.104082.

Bi, J., Lin, Y., Dong, Q., Yuan, H., Zhou, M., 2021. Large-scale water quality prediction with integrated deep neural network. Inf. Sci. 571, 191–205. https://doi.org/10.1016/j.ins.2021.04.057.

Bi, J., Zhang, L., Yuan, H., Zhang, J., 2023. Multi-indicator water quality prediction with attention-assisted bidirectional LSTM and encoder-decoder. Inf. Sci. 625, 65–80. https://doi.org/10.1016/j.ins.2022.12.091.
Blöschl, G., 2006. Rainfall-Runoff Modeling of Ungauged Catchments. In: Encyclopedia of Hydrological Sciences. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470848944.hsa140
Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: A review. Hydrol. Process. 9, 251–290. https://doi.org/10.1002/hyp.3360090305.
Braşoveanu, A.M.P., Andonie, R., 2020. Visualizing Transformers for NLP: A Brief Survey, in: 2020 24th International Conference Information Visualisation (IV). Presented at the 2020 24th International Conference Information Visualisation (IV), pp. 270–279. https://doi.org/10.1109/IV51561.2020.00051.
Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.
Brunner, P., Simmons, C.T., 2012. HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model. Groundwater 50, 170–176. https://doi.org/10.1111/j.1745-6584.2011.00882.x.
Brunner, G.W., 1996. HEC-RAS (River Analysis System). Presented at the North American Water and Environment Congress & Destructive Water, ASCE, pp. 3782–3787.
Bulti, D.T., Abebe, B.G., 2020. A review of flood modeling methods for urban pluvial flood application. Model. Earth Syst. Environ. 6, 1293–1302. https://doi.org/10.1007/s40808-020-00803-z.
Cai, H., Liu, S., Shi, H., Zhou, Z., Jiang, S., Babovic, V., 2022. Toward improved lumped groundwater level predictions at catchment scale: Mutual integration of water balance mechanism and deep learning method. J. Hydrol. 613, 128495 https://doi.org/10.1016/j.jhydrol.2022.128495.
Camps-Valls, G., Reichstein, M., Zhu, X., Tuia, D., 2020. ADVANCING DEEP LEARNING FOR EARTH SCIENCES: FROM HYBRID MODELING TO INTERPRETABILITY, in: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, pp. 3979–3982. https://doi.org/10.1109/IGARSS39084.2020.9323558.
Camps-Valls, G., Tuia, D., Zhu, X.X., Reichstein, M., 2021. Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing. John Wiley & Sons, Climate Science and Geosciences.
Castelvecchi, D., 2016. Can we open the black box of AI? Nat. News 538, 20. https://doi.org/10.1038/538020a.
Chandra, R., Saha, U., Mujumdar, P.P., 2015. Model and parameter uncertainty in IDF relationships under climate change. Adv. Water Resour. 79, 127–139. https://doi.org/10.1016/j.advwatres.2015.02.011.
Chang, L.-C., Wang, W.-H., Chang, F.-J., 2021. Explore training self-organizing map methods for clustering high-dimensional flood inundation maps. J. Hydrol. 595, 125655 https://doi.org/10.1016/j.jhydrol.2020.125655.
Chawla, I., Osuri, K.K., Mujumdar, P.P., Niyogi, D., 2018. Assessment of the Weather Research and Forecasting (WRF) model for simulation of extreme rainfall events in the upper Ganga Basin. Hydrol. Earth Syst. Sci. 22, 1095–1117. https://doi.org/10.5194/hess-22-1095-2018.
Chen, R., Wang, X., Zhang, W., Zhu, X., Li, A., Yang, C., 2019. A hybrid CNN-LSTM model for typhoon formation forecasting. GeoInformatica 23, 375–396. https://doi.org/10.1007/s10707-019-00355-0.
Cheng, L., AghaKouchak, A., Gilleland, E., Katz, R.W., 2014. Non-stationary extreme value analysis in a changing climate. Clim. Change 127, 353–369. https://doi.org/10.1007/s10584-014-1254-5.
Cheng, G., Zhou, P., Han, J., 2016. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. IEEE Trans. Geosci. Remote Sens. 54, 7405–7415. https://doi.org/10.1109/TGRS.2016.2601622.
K. Cho B. van Merrienboer C. Gulcehre D. Bahdanau F. Bougares H. Schwenk Y. Bengio Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation 2014 https://doi.org/10.48550/arXiv.1406.1078.
Chollet, F., 2021. Deep Learning with Python, Second Edition. Simon and Schuster.
Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling [WWW Document]. arXiv.org. https://doi.org/10.48550/arXiv.1412.3555.
Clark, S.R., 2022. Unravelling groundwater time series patterns: Visual analytics-aided deep learning in the Namoi region of Australia. Environ. Model. Softw. 149, 105295 https://doi.org/10.1016/j.envsoft.2022.105295.
Clark, M.P., Bierkens, M.F.P., Samaniego, L., Woods, R.A., Uijlenhoet, R., Bennett, K.E., Pauwels, V.R.N., Cai, X., Wood, A.W., Peters-Lidard, C.D., 2017. The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism. Hydrol. Earth Syst. Sci. 21, 3427–3440. https://doi.org/10.5194/hess-21-3427-2017.
Coles, S., Pericchi, L.R., Sisson, S., 2003. A fully probabilistic approach to extreme rainfall modeling. J. Hydrol. 273, 35–50. https://doi.org/10.1016/S0022-1694(02)00353-0.
Coulibaly, P., Baldwin, C.K., 2005. Nonstationary hydrological time series forecasting using nonlinear dynamic methods. J. Hydrol. 307, 164–174. https://doi.org/10.1016/j.jhydrol.2004.10.008.
Crawford, N.H., Linsley, R.K., 1966. DIGITAL SIMULATION IN HYDROLOGY' STANFORD WATERSHED MODEL 4.
Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., Bharath, A.A., 2018. Generative Adversarial Networks: An Overview. IEEE Signal Process. Mag. 35, 53–65. https://doi.org/10.1109/MSP.2017.2765202.
Dai, J.J., Wang, Yiheng, Qiu, X., Ding, D., Zhang, Y., Wang, Yanzhang, Jia, X., Zhang, C. L., Wan, Y., Li, Z., Wang, J., Huang, S., Wu, Z., Wang, Yang, Yang, Y., She, B., Shi, D., Lu, Q., Huang, K., Song, G., 2019. BigDL: A Distributed Deep Learning Framework for Big Data, in: Proceedings of the ACM Symposium on Cloud Computing, SoCC '19. Association for Computing Machinery, New York, NY, USA, pp. 50–60. https://doi.org/10.1145/3357223.3362707.
Daw, A., Karpatne, A., Watkins, W.D., Read, J.S., Kumar, V., 2022. Physics-Guided Neural Networks (PGNN): An Application in Lake Temperature Modeling. Knowledge-Guided Machine Learning. Chapman and Hall/CRC, in.
Deman, V., Koppa, A., Miralles, D., 2022. Seasonal Forecasting of Horn of Africa’s Long Rains Using Physics-Guided Machine Learning (No. EGU22-9719). Presented at the EGU22, Copernicus Meetings. https://doi.org/10.5194/egusphere-egu22-9719.
Dick, S., 2019. Artificial Intelligence. Harv. Data Sci. Rev. 1 https://doi.org/10.1162/99608f92.92fe150c.
Dikshit, A., Pradhan, B., 2021. Interpretable and explainable AI (XAI) model for spatial drought prediction. Sci. Total Environ. 801, 149797 https://doi.org/10.1016/j.scitotenv.2021.149797.
Dikshit, A., Pradhan, B., Alamri, A.M., 2021. Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model. Sci. Total Environ. 755, 142638 https://doi.org/10.1016/j.scitotenv.2020.142638.
Dikshit, A., Pradhan, B., Assiri, M.E., Almazroui, M., Park, H.-J., 2022. Solving transparency in drought forecasting using attention models. Sci. Total Environ. 837, 155856 https://doi.org/10.1016/j.scitotenv.2022.155856.
Ding, Y., Zhu, Y., Feng, J., Zhang, P., Cheng, Z., 2020. Interpretable spatio-temporal attention LSTM model for flood forecasting. Neurocomputing 403, 348–359. https://doi.org/10.1016/j.neucom.2020.04.110.
Donigian, A.S., Bicknell, B.R., Imhoff, J.C., 1995. Hydrological Simulation Program - Fortran (HSPF). Comput. Models Watershed Hydrol. 395–442.
Drucker, H., Wu, D., Vapnik, V.N., 1999. Support vector machines for spam categorization. IEEE Trans. Neural Netw. 10, 1048–1054. https://doi.org/10.1109/72.788645.
Du, S., Li, T., Yang, Y., Horng, S.-J., 2020. Multivariate time series forecasting via attention-based encoder–decoder framework. Neurocomputing 388, 269–279. https://doi.org/10.1016/j.neucom.2019.12.118.
Du, M., Liu, N., Yang, F., Ji, S., Hu, X., 2019. On Attribution of Recurrent Neural Network Predictions via Additive Decomposition. In: The World Wide Web Conference, WWW '19. Association for Computing Machinery, New York, NY, USA, pp. 383–393. https://doi.org/10.1145/3308558.3313545.
Elshorbagy, A., Corzo, G., Srinivasulu, S., Solomatine, D.P., 2010. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology. Hydrol. Earth Syst. Sci. 14, 1931–1941. https://doi.org/10.5194/hess-14-1931-2010.
Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., Piepmeier, J.R., Koster, R.D., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., Van Zyl, J., 2010. The Soil Moisture Active Passive (SMAP) Mission. IEEE.
Erion, G., Janizek, J.D., Sturmfels, P., Lundberg, S.M., Lee, S.-I., 2021. Improving performance of deep learning models with axiomatic attribution priors and expected gradients. Nat. Mach. Intell. 3, 620–631. https://doi.org/10.1038/s42256-021-00343-w.
Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. Oceans 99, 10143–10162. https://doi.org/10.1029/94JC00572.
Fang, K., Shen, C., 2020. Near-Real-Time Forecast of Satellite-Based Soil Moisture Using Long Short-Term Memory with an Adaptive Data Integration Kernel. J. Hydrometeorol. 21, 399–413. https://doi.org/10.1175/JHM-D-19-0169.1.
Fang, K., Shen, C., Kifer, D., Yang, X., 2017. Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental U.S. Using a Deep Learning Neural Network. Geophys. Res. Lett. 44, 11030–11039. https://doi.org/10.1002/2017GL075619.
Fang, Z., Wang, Y., Peng, L., Hong, H., 2021. Predicting flood susceptibility using LSTM neural networks. J. Hydrol. 594, 125734 https://doi.org/10.1016/j.jhydrol.2020.125734.
Fatichi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., Downer, C.W., Camporese, M., Davison, J.H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., Tarboton, D., 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. J. Hydrol. 537, 45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026.
Feng, D., Fang, K., Shen, C., 2020. Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales. Water Resour. Res. 56, e2019WR026793 https://doi.org/10.1029/2019WR026793.
Feng, C.-M., Yan, Y., Fu, H., Chen, L., Xu, Y., 2021. Task Transformer Network for Joint MRI Reconstruction and Super-Resolution. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (Eds.), Medical Image Computing and Computer Assisted Intervention –. Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 307–317. https://doi.org/10.1007/978-3-030-87231-1_30.
Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Stat. 29, 1189–1232.
Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, in. In: Proceedings of the 33rd International Conference on Machine Learning. Presented at the International Conference on Machine Learning, pp. 1050–1059.
Ganguly, A.R., Kodra, E.A., Agrawal, A., Banerjee, A., Boriah, S., Chatterjee, S.n., Chatterjee, S.o., Choudhary, A., Das, D., Faghmous, J., Ganguli, P., Ghosh, S., Hayhoe, K., Hays, C., Hendrix, W., Fu, Q., Kawale, J., Kumar, D., Kumar, V.,

Liao, W., Liess, S., Mawalagedara, R., Mithal, V., Oglesby, R., Salvi, K., Snyder, P.K., Steinhaeuser, K., Wang, D., Wuebbles, D., 2014. Toward enhanced understanding and projections of climate extremes using physics-guided data mining techniques. Nonlinear Process. Geophys. 21, 777–795. https://doi.org/10.5194/npg-21-777-2014.

Gavahi, K., Foroumandi, E., Moradkhani, H., 2023. A deep learning-based framework for multi-source precipitation fusion. Remote Sens. Environ. 295, 113723 https://doi.org/10.1016/j.rse.2023.113723.

Ghaith, M., Siam, A., Li, Z., El-Dakhakhni, W., 2020. Hybrid Hydrological Data-Driven Approach for Daily Streamflow Forecasting. J. Hydrol. Eng. 25, 04019063. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001866.

Giffard-Roisin, S., Yang, M., Charpiat, G., Kumler Bonfanti, C., Kégl, B., Monteleoni, C., 2020. Tropical Cyclone Track Forecasting Using Fused Deep Learning From Aligned Reanalysis Data. Front, Big Data, p. 3.

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. J. Comput. Graph. Stat. 24, 44–65. https://doi.org/10.1080/10618600.2014.907095.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative Adversarial Networks. https://doi.org/10.48550/arXiv.1406.2661.

Gowri, R., Dey, P., Mujumdar, P.P., 2021. A hydro-climatological outlook on the long-term availability of water resources in Cauvery river basin. Water Secur. 14, 100102 https://doi.org/10.1016/j.wasec.2021.100102.

Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw., IJCNN 2005 18, 602–610. https://doi.org/10.1016/j.neunet.2005.06.042.

Gulli, A., Pal, S., 2017. Deep Learning with Keras. Packt Publishing Ltd.

Guzman, S.M., Paz, J.O., Tagert, M.L.M., 2017. The Use of NARX Neural Networks to Forecast Daily Groundwater Levels. Water Resour. Manag. 31, 1591–1603. https://doi.org/10.1007/s11269-017-1598-5.

Ham, Y.-G., Kim, J.-H., Luo, J.-J., 2019. Deep learning for multi-year ENSO forecasts. Nature 573, 568–572. https://doi.org/10.1038/s41586-019-1559-7.

Han, H., Morrison, R.R., 2022. Improved runoff forecasting performance through error predictions using a deep-learning approach. J. Hydrol. 608, 127653 https://doi.org/10.1016/j.jhydrol.2022.127653.

Hastie, T., Tibshirani, R., Friedman, J., 2009. Unsupervised Learning, in: Hastie, T., Tibshirani, R., Friedman, J. (Eds.), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Series in Statistics. Springer, New York, NY, pp. 485–585. https://doi.org/10.1007/978-0-387-84858-7_14.

He, X., Tan, E.-L., Bi, H., Zhang, X., Zhao, S., Lei, B., 2022. Fully transformer network for skin lesion analysis. Med. Image Anal. 77, 102357 https://doi.org/10.1016/j.media.2022.102357.

He, T., Wang, N., Zhang, D., 2021. Theory-guided full convolutional neural network: An efficient surrogate model for inverse problems in subsurface contaminant transport. Adv. Water Resour. 157, 104051 https://doi.org/10.1016/j.advwatres.2021.104051.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR 770–778. https://doi.org/10.1109/CVPR.2016.90.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the Dimensionality of Data with Neural Networks. Science 313, 504–507. https://doi.org/10.1126/science.1127647.

Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors No. arXiv:1207.0580, arXiv. https://doi.org/10.48550/arXiv.1207.0580.

Hochreiter, S., 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 6, 107–116. https://doi.org/10.1142/S0218488598000094.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Hoerl, A.E., Kennard, R.W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 12, 55–67. https://doi.org/10.2307/1267351.

Hu, P., Perazzi, F., Heilbron, F.C., Wang, O., Lin, Z., Saenko, K., Sclaroff, S., 2021. Real-Time Semantic Segmentation With Fast Attention. IEEE Robot. Autom. Lett. 6, 263–270. https://doi.org/10.1109/LRA.2020.3039744.

Huang, Z., Dumitru, C.O., Ren, J., 2021. Physics-Aware Feature Learning of Sar Images with Deep Neural Networks: A Case Study, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. Presented at the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 1264–1267. https://doi.org/10.1109/IGARSS47720.2021.9554842.

Iliadis, L.S., Maris, F., 2007. An Artificial Neural Network model for mountainous water-resources management: The case of Cyprus mountainous watersheds. Environ. Model. Softw. 22, 1066–1072. https://doi.org/10.1016/j.envsoft.2006.05.026.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in: Proceedings of the 32nd International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 448–456.

Iwana, B.K., Uchida, S., 2021. An empirical survey of data augmentation for time series classification with neural networks. PLOS ONE 16, e0254841.

Jain, A., Kumar Varshney, A., Chandra Joshi, U., 2001. Short-Term Water Demand Forecast Modelling at IIT Kanpur Using Artificial Neural Networks. Water Resour. Manag. 15, 299–321. https://doi.org/10.1023/A:1014415503476.

Jeppesen, J.H., Jacobsen, R.H., Inceoglu, F., Toftegaard, T.S., 2019. A cloud detection algorithm for satellite imagery based on deep learning. Remote Sens. Environ. 229, 247–259. https://doi.org/10.1016/j.rse.2019.03.039.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional Architecture for Fast Feature Embedding, in: Proceedings of the 22nd ACM International Conference on Multimedia, MM '14. Association for Computing Machinery, New York, NY, USA, pp. 675–678. https://doi.org/10.1145/2647868.2654889.

Jia, X., Willard, J., Karpatne, A., Read, J.S., Zwart, J.A., Steinbach, M., Kumar, V., 2021. Physics-Guided Machine Learning for Scientific Discovery: An Application in Simulating Lake Temperature Profiles. ACMIMS Trans. Data Sci. 2, 20: 1 –20, 26. https://doi.org/10.1145/3447814.

Jiang, W., 2018. Object-based deep convolutional autoencoders for high-resolution remote sensing image classification. J. Appl. Remote Sens. 12, 035002 https://doi.org/10.1117/1.JRS.12.035002.

Jiang, S., Zheng, Y., Solomatine, D., 2020. Improving AI System Awareness of Geoscience Knowledge: Symbiotic Integration of Physical Approaches and Deep Learning. Geophys. Res. Lett. 47, e2020GL088229 https://doi.org/10.1029/2020GL088229.

Jiang, S., Zheng, Y., Wang, C., Babovic, V., 2022. Uncovering Flooding Mechanisms Across the Contiguous United States Through Interpretive Deep Learning on Representative Catchments. Water Resour. Res. 58, e2021WR030185 https://doi.org/10.1029/2021WR030185.

Kao, I.-F., Zhou, Y., Chang, L.-C., Chang, F.-J., 2020. Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. J. Hydrol. 583, 124631 https://doi.org/10.1016/j.jhydrol.2020.124631.

Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. IEEE Trans. Knowl. Data Eng. 29, 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168.

Karthikeyan, L., Mishra, A.K., 2021. Multi-layer high-resolution soil moisture estimation using machine learning over the United States. Remote Sens. Environ. 266, 112706 https://doi.org/10.1016/j.rse.2021.112706.

Kaur, S., Gupta, S., Singh, S., Koundal, D., Zaguia, A., 2022. Convolutional neural network based hurricane damage detection using satellite images. Soft Comput. 26, 7831–7845. https://doi.org/10.1007/s00500-022-06805-6.

Kendall, A., Gal, Y., 2017. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? Advances in Neural Information Processing Systems. Curran Associates Inc.

Kerem Cigizoglu, H., Kisi, Ö., 2006. Methods to improve the neural network performance in suspended sediment estimation. J. Hydrol. 317, 221–238. https://doi.org/10.1016/j.jhydrol.2005.05.019.

A. Khandelwal S. Xu X. Li X. Jia M. Stienbach C. Duffy J. Nieber V. Kumar Physics Guided Machine Learning Methods for Hydrology 2020 https://doi.org/10.48550/arXiv.2012.02854.

Kim, S., Kim, H., Lee, J., Yoon, S., Kahou, S.E., Kashinath, K., Prabhat, M., 2019. Deep-Hurricane-Tracker: Tracking and Forecasting Extreme Climate Events, in: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). Presented at the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1761–1769. https://doi.org/10.1109/WACV.2019.00192.

Kingston, G.B., Lambert, M.F., Maier, H.R., 2005. Bayesian training of artificial neural networks used for water resources modeling. Water Resour. Res. 41 https://doi.org/10.1029/2005WR004152.

Klotz, D., Kratzert, F., Gauch, M., Keefe Sampson, A., Brandstetter, J., Klambauer, G., Hochreiter, S., Nearing, G., 2022. Uncertainty estimation with deep learning for rainfall–runoff modeling. Hydrol. Earth Syst. Sci. 26, 1673–1693. https://doi.org/10.5194/hess-26-1673-2022.

N. Kokhlikyan V. Miglani M. Martin E. Wang B. Alsallakh J. Reynolds A. Melnikov N. Kliushkina C. Araya S. Yan O. Reblitz-Richardson Captum: A unified and generic model interpretability library for PyTorch 2020 https://doi.org/10.48550/arXiv.2009.07896.

Konapala, G., Mishra, A., 2020. Quantifying Climate and Catchment Control on Hydrological Drought in the Continental United States. Water Resour. Res. 56, e2018WR024620 https://doi.org/10.1029/2018WR024620.

Kotsiantis, S.B., Zaharakis, I.D., Pintelas, P.E., 2006. Machine learning: a review of classification and combining techniques. Artif. Intell. Rev. 26, 159–190. https://doi.org/10.1007/s10462-007-9052-3.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M., 2018. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrol. Earth Syst. Sci. 22, 6005–6022. https://doi.org/10.5194/hess-22-6005-2018.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A.K., Hochreiter, S., Nearing, G.S., 2019. Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning. Water Resour. Res. 55, 11344–11354. https://doi.org/10.1029/2019WR026065.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems. Curran Associates Inc.

J. Kukačka V. Golkov D. Cremers Regularization for Deep Learning: A Taxonomy 2017 https://doi.org/10.48550/arXiv.1710.10686.

Kusiak, A., 2017. Smart manufacturing must embrace big data. Nature 544, 23–25. https://doi.org/10.1038/544023a.

Lallahem, S., Mania, J., Hani, A., Najjar, Y., 2005. On the use of neural networks to evaluate groundwater levels in fractured media. J. Hydrol. 307, 92–111. https://doi.org/10.1016/j.jhydrol.2004.10.005.

Laloy, E., Hérault, R., Lee, J., Jacques, D., Linde, N., 2017. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. Adv. Water Resour. 110, 387–405. https://doi.org/10.1016/j.advwatres.2017.09.029.

Laloy, E., Hérault, R., Jacques, D., Linde, N., 2018. Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network. Water Resour. Res. 54, 381–406. https://doi.org/10.1002/2017WR022148.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324. https://doi.org/10.1109/5.726791.

Li, Q., Zhu, Y., Shangguan, W., Wang, X., Li, L., Yu, F., 2022. An attention-aware LSTM model for soil moisture and soil temperature prediction. Geoderma 409, 115651. https://doi.org/10.1016/j.geoderma.2021.115651.

Li, W., Liu, K., Zhang, L., Cheng, F., 2020. Object detection based on an adaptive attention mechanism. Sci. Rep. 10, 11307. https://doi.org/10.1038/s41598-020-67529-x.

Li, W., Kiaghadi, A., Dawson, C., 2021. Exploring the best sequence LSTM modeling architecture for flood prediction. Neural Comput. Appl. 33, 5571–5580. https://doi.org/10.1007/s00521-020-05334-3.

Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J. Geophys. Res. Atmospheres 99, 14415–14428. https://doi.org/10.1029/94JD00483.

Liang, X., Lettenmaier, D.P., Wood, E.F., 1996. One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model. J. Geophys. Res. Atmospheres 101, 21403–21422. https://doi.org/10.1029/96JD01448.

Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resour. Res. 43 https://doi.org/10.1029/2006WR005756.

Liu, D., Mishra, A.K., Yu, Z., Lü, H., Li, Y., 2021. Support vector machine and data assimilation framework for Groundwater Level Forecasting using GRACE satellite data. J. Hydrol. 603, 126929 https://doi.org/10.1016/j.jhydrol.2021.126929.

Liu, J., Rahmani, F., Lawson, K., Shen, C., 2022a. A Multiscale Deep Learning Model for Soil Moisture Integrating Satellite and In Situ Data. Geophys. Res. Lett. 49, e2021GL096847 https://doi.org/10.1029/2021GL096847.

Liu, B., Zou, X., Yi, S., Sneeuw, N., Li, J., Cai, J., 2022b. Reconstructing GRACE-like time series of high mountain glacier mass anomalies. Remote Sens. Environ. 280, 113177. https://doi.org/10.1016/j.rse.2022.113177.

S. Lundberg S.-I. Lee A Unified Approach to Interpreting Model Predictions 2017 https://doi.org/10.48550/arXiv.1705.07874.

Máchová, R., Hub, M., Lnenicka, M., 2018. Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective. Aslib J. Inf. Manag. 70, 252–268. https://doi.org/10.1108/AJIM-02-2018-0026.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. Proc. Fifth Berkeley Symp. Math. Stat. Probab. Vol. 1 Stat. 5.1, 281–298.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. IEEE Trans. Geosci. Remote Sens. 55, 645–657. https://doi.org/10.1109/TGRS.2016.2612821.

Maier, H.R., Dandy, G.C., 1996. The Use of Artificial Neural Networks for the Prediction of Water Quality Parameters. Water Resour. Res. 32, 1013–1022. https://doi.org/10.1029/96WR03529.

Maqsood, I., Khan, M.R., Abraham, A., 2004. An ensemble of neural networks for weather forecasting. Neural Comput. Appl. 13, 112–122. https://doi.org/10.1007/s00521-004-0413-4.

Martínez-Fernández, J., Ceballos, A., 2005. Mean soil moisture estimation using temporal stability analysis. J. Hydrol. 312, 28–38. https://doi.org/10.1016/j.jhydrol.2005.02.007.

Mas, J.F., Flores, J.J., 2008. The application of artificial neural networks to the analysis of remotely sensed data. Int. J. Remote Sens. 29, 617–663. https://doi.org/10.1080/01431160701352154.

Mirza, M., Osindero, S., 2014. Conditional Generative Adversarial Nets (No. arXiv: 1411.1784). arXiv. https://doi.org/10.48550/arXiv.1411.1784.

Mishra, A., Alnahit, A., Campbell, B., 2021. Impact of land uses, drought, flood, wildfire, and cascading events on water quality and microbial communities: A review and analysis. J. Hydrol. 596, 125707 https://doi.org/10.1016/j.jhydrol.2020.125707.

Mishra, A.K., Desai, V.R., 2006. Drought forecasting using feed-forward recursive neural network. Ecol. Model. 198, 127–138. https://doi.org/10.1016/j.ecolmodel.2006.04.017.

Mishra, A., Desai, V., Singh, V., 2007. Drought Forecasting Using a Hybrid Stochastic and Neural Network Model. J. Hydrol. Eng. - J HYDROL ENG 12. doi: 10.1061/(ASCE)1084-0699(2007)12: 6 (626).

Mishra, A., Mukherjee, S., Merz, B., Singh, V.P., Wright, D.B., Villarini, G., Paul, S., Kumar, D.N., Khedun, C.P., Niyogi, D., Schumann, G., Stedinger, J.R., 2022. An Overview of Flood Concepts, Challenges, and Future Directions. J. Hydrol. Eng. 27, 03122001. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002164.

Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J. Hydrol. 391, 202–216. https://doi.org/10.1016/j.jhydrol.2010.07.012.

Mo, S., Zabaras, N., Shi, X., Wu, J., 2019. Deep Autoregressive Neural Networks for High-Dimensional Inverse Problems in Groundwater Contaminant Source Identification. Water Resour. Res. 55, 3856–3881. https://doi.org/10.1029/2018WR024638.

Mokhtar, A., Jalali, M., He, H., Al-Ansari, N., Elbeltagi, A., Alsafadi, K., Abdo, H.G., Sammen, S.S., Gyasi-Agyei, Y., Rodrigo-Comino, J., 2021. Estimation of SPEI Meteorological Drought Using Machine Learning Algorithms. IEEE Access 9, 65503–65523. https://doi.org/10.1109/ACCESS.2021.3074305.

Mujumdar, P.P., Kumar, D.N., 2012. Floods in a Changing Climate: Hydrologic Modeling. Cambridge University Press.

Mukherjee, S., Mishra, A.K., 2021. Increase in Compound Drought and Heatwaves in a Warming World. Geophys. Res. Lett. 48, e2020GL090617 https://doi.org/10.1029/2020GL090617.

Murakami, H., Delworth, T.L., Cooke, W.F., Kapnick, S.B., Hsu, P.-C., 2022. Increasing Frequency of Anomalous Precipitation Events in Japan Detected by a Deep Learning Autoencoder. Earths. Future 10, e2021EF002481. https://doi.org/10.1029/2021EF002481.

Niu, G.-Y., Yang, Z.-L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., Xia, Y., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. J. Geophys. Res. Atmospheres 116. https://doi.org/10.1029/2010JD015139.

Nourani, V., Baghanam, A.H., Adamowski, J., Gebremichael, M., 2013. Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling. J. Hydrol. 476, 228–243. https://doi.org/10.1016/j.jhydrol.2012.10.054.

Oguz, E.A., Depina, I., Myhre, B., Devoli, G., Rustad, H., Thakur, V., 2022. IoT-based hydrological monitoring of water-induced landslides: a case study in central Norway. Bull. Eng. Geol. Environ. 81, 217. https://doi.org/10.1007/s10064-022-02721-z.

Oleson, K.W., Niu, G.-Y., Yang, Z.-L., Lawrence, D.M., Thornton, P.E., Lawrence, P.J., Stöckli, R., Dickinson, R.E., Bonan, G.B., Levis, S., Dai, A., Qian, T., 2008. Improvements to the Community Land Model and their impact on the hydrological cycle. J. Geophys. Res. Biogeosciences 113. https://doi.org/10.1029/2007JG000563.

Oosthuizen, R.M., 2022. The Fourth Industrial Revolution – Smart Technology, Artificial Intelligence, Robotics and Algorithms: Industrial Psychologists in Future Workplaces. Front. Artif, Intell, p. 5.

Pan, B., Hsu, K., AghaKouchak, A., Sorooshian, S., 2019. Improving Precipitation Estimation Using Convolutional Neural Network. Water Resour. Res. 55, 2301–2321. https://doi.org/10.1029/2018WR024090.

Pan, S.J., Yang, Q., 2010. A Survey on Transfer Learning. IEEE Trans. Knowl. Data Eng. 22, 1345–1359. https://doi.org/10.1109/TKDE.2009.191.

Pan, M., Zhou, H., Cao, J., Liu, Y., Hao, J., Li, S., Chen, C.-H., 2020. Water Level Prediction Model Based on GRU and CNN. IEEE Access 8, 60090–60100. https://doi.org/10.1109/ACCESS.2020.2982433.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in PyTorch.

Paul, S., Nagesh Kumar, D., 2018. Spectral-spatial classification of hyperspectral data with mutual information based segmented stacked autoencoder approach. ISPRS J. Photogramm. Remote Sens. 138, 265–280. https://doi.org/10.1016/j.isprsjprs.2018.02.001.

Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. Lond. Edinb. Dublin Philos. Mag. J. Sci. 2, 559–572. https://doi.org/10.1080/14786440109462720.

Pinder, G.F., Gray, W.G., 2013. Finite Element Simulation in Surface and Subsurface Hydrology. Elsevier.

Prasad, D.V.V., Venkataramana, L.Y., Kumar, P.S., Prasannamedha, G., Harshana, S., Srividya, S.J., Harrinei, K., Indraganti, S., 2022. Analysis and prediction of water quality using deep learning and auto deep learning techniques. Sci. Total Environ. 821, 153311 https://doi.org/10.1016/j.scitotenv.2022.153311.

A. Radford L. Metz S. Chintala Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks 2016 https://doi.org/10.48550/arXiv.1511.06434.

Raghavendra. N, S., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. Appl. Soft Comput. 19, 372–386. https://doi.org/10.1016/j.asoc.2014.02.002.

Rahman, A.T.M.S., Hosono, T., Quilty, J.M., Das, J., Basak, A., 2020. Multiscale groundwater level forecasting: Coupling new machine learning approaches with wavelet transforms. Adv. Water Resour. 141, 103595 https://doi.org/10.1016/j.advwatres.2020.103595.

Rasp, S., Pritchard, M.S., Gentine, P., 2018. Deep learning to represent subgrid processes in climate models. Proc. Natl. Acad. Sci. 115, 9684–9689. https://doi.org/10.1073/pnas.1810286115.

Read, J.S., Jia, X., Willard, J., Appling, A.P., Zwart, J.A., Oliver, S.K., Karpatne, A., Hansen, G.J.A., Hanson, P.C., Watkins, W., Steinbach, M., Kumar, V., 2019. Process-Guided Deep Learning Predictions of Lake Water Temperature. Water Resour. Res. 55, 9173–9190. https://doi.org/10.1029/2019WR024922.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Rezaie-balf, M., Naganna, S.R., Ghaemi, A., Deka, P.C., 2017. Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting. J. Hydrol. 553, 356–373. https://doi.org/10.1016/j.jhydrol.2017.08.006.

M.T. Ribeiro S. Singh C. Guestrin "Why Should I Trust You?": Explaining the Predictions of Any Classifier 2016 https://doi.org/10.48550/arXiv.1602.04938.

Rosecrans, C.Z., Nolan, B.T., Gronberg, J.M., 2017. Prediction and visualization of redox conditions in the groundwater of Central Valley. California. J. Hydrol. 546, 341–356. https://doi.org/10.1016/j.jhydrol.2017.01.014.
Rubin, A.G., Carlin, J.B., Stern, H.S., Donald, B., 1995. Bayesian Data Analysis. Chapman and Hall/CRC, New York. https://doi.org/10.1201/9780429258411.
Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1985. Learning Internal Representations by Error Propagation. CALIFORNIA UNIV SAN DIEGO LA JOLLA INST FOR COGNITIVE SCIENCE.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. https://doi.org/10.1038/323533a0.
Sadler, J.M., Appling, A.P., Read, J.S., Oliver, S.K., Jia, X., Zwart, J.A., Kumar, V., 2022. Multi-Task Deep Learning of Daily Streamflow and Water Temperature. Water Resour. Res. 58, e2021WR030138 https://doi.org/10.1029/2021WR030138.
Saha, S., Gayen, A., Bayen, B., 2022. Deep learning algorithms to develop Flood susceptibility map in Data-Scarce and Ungauged River Basin in India. Stoch. Environ. Res. Risk Assess. https://doi.org/10.1007/s00477-022-02195-1.
Scharfferberg, W.A., Fleming, M.J., 2006. Hydrologic modeling system HEC-HMS: user's manual. US Army Corps Eng. Hydrol. Eng, Cent, p. 2006.
Schoppa, L., Disse, M., Bachmair, S., 2020. Evaluating the performance of random forest for large-scale flood discharge simulation. J. Hydrol. 590, 125531 https://doi.org/10.1016/j.jhydrol.2020.125531.
Sermet, Y., Villanueva, P., Sit, M.A., Demir, I., 2020. Crowdsourced approaches for stage measurements at ungauged locations using smartphones. Hydrol. Sci. J. 65, 813–822. https://doi.org/10.1080/02626667.2019.1659508.
Shabri, A., Suhartono, 2012. Streamflow forecasting using least-squares support vector machines. Hydrol. Sci. J. 57, 1275–1293. https://doi.org/10.1080/02626667.2012.714468.
Shen, C., 2018. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. Water Resour. Res. 54, 8558–8593. https://doi.org/10.1029/2018WR022643.
Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.-J., Ganguly, S., Hsu, K.-L., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X., Tsai, W.-P., 2018. HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community. Hydrol. Earth Syst. Sci. 22, 5639–5656. https://doi.org/10.5194/hess-22-5639-2018.
Shen, C., Chen, X., Laloy, E., 2021. Editorial: Broadening the Use of Machine Learning in Hydrology. Front, Water, p. 3.
Shigidi, A., Garcia, L.A., 2003. Parameter Estimation in Groundwater Hydrology Using Artificial Neural Networks. J. Comput. Civ. Eng. 17, 281–289. https://doi.org/10.1061/(ASCE)0887-3801(2003)17:4(281).
Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. J. Big Data 6, 60. https://doi.org/10.1186/s40537-019-0197-0.
Sikorska, A.E., Montanari, A., Koutsoyiannis, D., 2015. Estimating the Uncertainty of Hydrological Predictions through Data-Driven Resampling Techniques. J. Hydrol. Eng. 20, A4014009. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000926.
Simard, P., Steinkraus, D., Platt, J., 2003. Best Practices for Convolutional Neural Networks.
K. Simonyan A. Zisserman Very Deep Convolutional Networks for Large-Scale Image Recognition 2015 https://doi.org/10.48550/arXiv.1409.1556.
K. Simonyan A. Vedaldi A. Zisserman Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps 2014 https://doi.org/10.48550/arXiv.1312.6034.
Singh, K.P., Basant, A., Malik, A., Jain, G., 2009. Artificial neural network modeling of the river water quality—A case study. Ecol. Model. 220, 888–895. https://doi.org/10.1016/j.ecolmodel.2009.01.004.
Singh, V.P., Woolhiser, D.A., 2002. Mathematical Modeling of Watershed Hydrology. J. Hydrol. Eng. 7, 270–292. https://doi.org/10.1061/(ASCE)1084-0699(2002)7:4(270).
Sit, M., Demiray, B.Z., Xiang, Z., Ewing, G.J., Sermet, Y., Demir, I., 2020. A comprehensive review of deep learning applications in hydrology and water resources. Water Sci. Technol. 82, 2635–2670. https://doi.org/10.2166/wst.2020.369.
Sivapalan, M., Blöschl, G., Zhang, L., Vertessy, R., 2003. Downward approach to hydrological prediction. Hydrol. Process. 17, 2101–2111. https://doi.org/10.1002/hyp.1425.
Skamarock, W.C., Klemp, J., Dudhia, J., Gill, D.O., Barker, D., Wang, W., Powers, J.G., 2008. A Description of the Advanced Research WRF Version 3 (27), 3–27.
J.T. Springenberg A. Dosovitskiy T. Brox M. Riedmiller Striving for Simplicity: The All Convolutional Net 2015 https://doi.org/10.48550/arXiv.1412.6806.
Sreeparvathy, V., Srinivas, V.V., 2022. Meteorological flash droughts risk projections based on CMIP6 climate change scenarios. Npj Clim. Atmospheric Sci. 5, 1–17. https://doi.org/10.1038/s41612-022-00302-1.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.
Subramanian, M., Shanmugavadivel, K., Nandhini, P.S., 2022. On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves. Neural Comput. Appl. 34, 13951–13968. https://doi.org/10.1007/s00521-022-07246-w.
Sun, J., Hu, L., Li, D., Sun, K., Yang, Z., 2022. Data-driven models for accurate groundwater level prediction and their practical significance in groundwater management. J. Hydrol. 608, 127630 https://doi.org/10.1016/j.jhydrol.2022.127630.
Sundararajan, M., Taly, A., Yan, Q., 2017. Axiomatic Attribution for Deep Networks, in: Proceedings of the 34th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 3319–3328.
Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to Sequence Learning with Neural Networks. Advances in Neural Information Processing Systems. Curran Associates Inc.
Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning, second edition. MIT Press, An Introduction.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going Deeper with Convolutions. https://doi.org/10.48550/arXiv.1409.4842.
Tang, X., Tu, Z., Wang, Y., Liu, M., Li, D., Fan, X., 2022. Automatic Detection of Coseismic Landslides Using a New Transformer Method. Remote Sens. 14, 2884. https://doi.org/10.3390/rs14122884.
Tartakovsky, A.M., Marrero, C.O., Perdikaris, P., Tartakovsky, G.D., Barajas-Solano, D., 2020. Physics-Informed Deep Neural Networks for Learning Parameters and Constitutive Relationships in Subsurface Flow Problems. Water Resour. Res. 56, e2019WR026731 https://doi.org/10.1029/2019WR026731.
Tatem, A.J., Lewis, H.G., Atkinson, P.M., Nixon, M.S., 2001. Super-resolution target identification from remotely sensed images using a Hopfield neural network. IEEE Trans. Geosci. Remote Sens. 39, 781–796. https://doi.org/10.1109/36.917895.
Taylor, J.W., Buizza, R., 2002. Neural network load forecasting with weather ensemble predictions. IEEE Trans. Power Syst. 17, 626–632. https://doi.org/10.1109/TPWRS.2002.800906.
The Theano Development Team, Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., Ballas, N., Bastien, F., Bayer, J., Belikov, A., Belopolsky, A., Bengio, Y., Bergeron, A., Bergstra, J., Bisson, V., Bleecher Snyder, J., Bouchard, N., Boulanger-Lewandowski, N., Bouthillier, X., de Brébisson, A., Breuleux, O., Carrier, P.-L., Cho, K., Chorowski, J., Christiano, P., Cooijmans, T., Côté, M.-A., Côté, M., Courville, A., Dauphin, Y.N., Delalleau, O., Demouth, J., Desjardins, G., Dieleman, S., Dinh, L., Ducoffe, M., Dumoulin, V., Ebrahimi Kahou, S., Erhan, D., Fan, Z., Firat, O., Germain, M., Glorot, X., Goodfellow, I., Graham, M., Gulcehre, C., Hamel, P., Harlouchet, I., Heng, J.-P., Hidasi, B., Honari, S., Jain, A., Jean, S., Jia, K., Korobov, M., Kulkarni, V., Lamb, A., Lamblin, P., Larsen, E., Laurent, C., Lee, S., Lefrancois, S., Lemieux, S., Léonard, N., Lin, Z., Livezey, J.A., Lorenz, C., Lowin, J., Ma, Q., Manzagol, P.-A., Mastropietro, O., McGibbon, R.T., Memisevic, R., van Merriënboer, B., Michalski, V., Mirza, M., Orlandi, A., Pal, C., Pascanu, R., Pezeshki, M., Raffel, C., Renshaw, D., Rocklin, M., Romero, A., Roth, M., Sadowski, P., Salvatier, J., Savard, F., Schlüter, J., Schulman, J., Schwartz, G., Vlad Serban, I., Serdyuk, D., Shabanian, S., Simon, É., Spieckermann, S., Ramana Subramanyam, S., Sygnowski, J., Tanguay, J., van Tulder, G., Turian, J., Urban, S., Vincent, P., Visin, F., de Vries, H., Warde-Farley, D., Webb, D.J., Willson, M., Xu, K., Xue, L., Yao, L., Zhang, S., Zhang, Y., 2016. Theano: A Python framework for fast computation of mathematical expressions, arXiv e-prints.
Thompson, N.C., Greenewald, K., Lee, K., Manso, G.F., 2022. The Computational Limits of Deep Learning.
Tian, Y., Zhang, Y., 2022. A comprehensive survey on regularization strategies in machine learning. Inf. Fusion 80, 146–166. https://doi.org/10.1016/j.inffus.2021.11.005.
Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. J. R. Stat. Soc. Ser. B Methodol. 58, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
Tripathi, S., Srinivas, V.V., Nanjundiah, R.S., 2006. Downscaling of precipitation for climate change scenarios: A support vector machine approach. J. Hydrol. 330, 621–640. https://doi.org/10.1016/j.jhydrol.2006.04.030.
Tripathy, K.P., Mukherjee, S., Mishra, A.K., Mann, M.E., Williams, A.P., 2023. Climate change will accelerate the high-end risk of compound drought and heatwave events. Proc. Natl. Acad. Sci. 120, e2219825120 https://doi.org/10.1073/pnas.2219825120.
Troy, T.J., Wood, E.F., Sheffield, J., 2008. An efficient calibration method for continental-scale land surface modeling. Water Resour. Res. 44 https://doi.org/10.1029/2007WR006513.
Tu, T., Ishida, K., Ercan, A., Kiyama, M., Amagasaki, M., Zhao, T., 2021. Hybrid precipitation downscaling over coastal watersheds in Japan using WRF and CNN. J. Hydrol. Reg. Stud. 37, 100921 https://doi.org/10.1016/j.ejrh.2021.100921.
Vapnik, V., 1999. The Nature of Statistical Learning Theory. Springer Science & Business Media.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is All you Need. Advances in Neural Information Processing Systems. Curran Associates Inc.
Veettil, A.V., Mishra, A.K., Green, T.R., 2022. Explaining water security indicators using hydrologic and agricultural systems models. J. Hydrol. 607, 127463 https://doi.org/10.1016/j.jhydrol.2022.127463.
Vrugt, J.A., 2016. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. Environ. Model. Softw. 75, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013.
Wan, H., Xu, R., Zhang, M., Cai, Y., Li, J., Shen, X., 2022. A novel model for water quality prediction caused by non-point sources pollution based on deep learning and feature extraction methods. J. Hydrol. 612, 128081 https://doi.org/10.1016/j.jhydrol.2022.128081.
Wang, R., Yu, R., 2022. Physics-Guided Deep Learning for Dynamical Systems: A Survey.
Wang, J., Fisher, B.L., Wolff, D.B., 2008. Estimating Rain Rates from Tipping-Bucket Rain Gauge Measurements. J. Atmospheric Ocean. Technol. 25, 43–56. https://doi.org/10.1175/2007JTECHA895.1.
Wang, Q., Huang, J., Liu, R., Men, C., Guo, L., Miao, Y., Jiao, L., Wang, Y., Muhammad, S., Xia, X., 2020b. Sequence-based statistical downscaling and its

application to hydrologic simulations based on machine learning and big data. J. Hydrol. 586, 124875 https://doi.org/10.1016/j.jhydrol.2020.124875.
Wang, X., Kadıoğlu, S., 2021. Modeling uncertainty to improve personalized recommendations via Bayesian deep learning. J. Data Sci. Anal, Int, 10.1007/s41060-020-00241-1.
Wang, S., Peng, H., Liang, S., 2022. Prediction of estuarine water quality using interpretable machine learning approach. J. Hydrol. 605, 127320 https://doi.org/10.1016/j.jhydrol.2021.127320.
Wang, F., Tian, D., Lowe, L., Kalin, L., Lehrter, J., 2021. Deep Learning for Daily Precipitation and Temperature Downscaling. Water Resour. Res. 57, e2020WR029308 https://doi.org/10.1029/2020WR029308.
Wang, N., Zhang, D., Chang, H., Li, H., 2020a. Deep learning of subsurface flow via theory-guided neural network. J. Hydrol. 584, 124700 https://doi.org/10.1016/j.jhydrol.2020.124700.
Wei, M., You, X., 2022. Monthly rainfall forecasting by a hybrid neural network of discrete wavelet transformation and deep learning. Water Resour. Manag. 36, 4003–4018. https://doi.org/10.1007/s11269-022-03218-w.
Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. J. Big Data 3, 9. https://doi.org/10.1186/s40537-016-0043-6.
Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X., Xu, H., 2021. Time Series Data Augmentation for Deep Learning: A Survey, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. pp. 4653–4660. https://doi.org/10.24963/ijcai.2021/631.
Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., 2009. Niches, models, and climate change: Assessing the assumptions and uncertainties. Proc. Natl. Acad. Sci. 106, 19729–19736. https://doi.org/10.1073/pnas.0901639106.
Willard, J.D., Read, J.S., Appling, A.P., Oliver, S.K., Jia, X., Kumar, V., 2021. Predicting Water Temperature Dynamics of Unmonitored Lakes With Meta-Transfer Learning. Water Resour. Res. 57, e2021WR029579 https://doi.org/10.1029/2021WR029579.
Wood, E.F., Roundy, J.K., Troy, T.J., van Beek, L.P.H., Bierkens, M.F.P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P.R., Kollet, S., Lehner, B., Lettenmaier, D.P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., Whitehead, P., 2011. Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. Water Resour. Res. 47 https://doi.org/10.1029/2010WR010090.
Woods, R., Sivapalan, M., 1999. A synthesis of space-time variability in storm response: Rainfall, runoff generation, and routing. Water Resour. Res. 35, 2469–2485. https://doi.org/10.1029/1999WR900014.
Wu, N., Green, B., Ben, X., O'Banion, S., 2020. Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case.
Wu, H., Fang, W.-Z., Kang, Q., Tao, W.-Q., Qiao, R., 2019. Predicting Effective Diffusivity of Porous Media from Images by Deep Learning. Sci. Rep. 9, 20387. https://doi.org/10.1038/s41598-019-56309-x.
wu, Y., Liu, Z., Xu, W., Feng, J., Palaiahnakote, S., Lu, T., 2018. Context-Aware Attention LSTM Network for Flood Prediction, in: 2018 24th International Conference on Pattern Recognition (ICPR). Presented at the 2018 24th International Conference on Pattern Recognition (ICPR), pp. 1301–1306. https://doi.org/10.1109/ICPR.2018.8545385.
Xiang, Z., Yan, J., Demir, I., 2020. A Rainfall-Runoff Model With LSTM-Based Sequence-to-Sequence Learning. Water Resour. Res. 56, e2019WR025326 https://doi.org/10.1029/2019WR025326.
Xie, K., Liu, P., Zhang, J., Han, D., Wang, G., Shen, C., 2021. Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. J. Hydrol. 603, 127043 https://doi.org/10.1016/j.jhydrol.2021.127043.
Xing, H., Meng, Y., Wang, Z., Fan, K., Hou, D., 2018. Exploring geo-tagged photos for land cover validation with deep learning. ISPRS J. Photogramm. Remote Sens. 141, 237–251. https://doi.org/10.1016/j.isprsjprs.2018.04.025.
Xu, J., Fan, H., Luo, M., Li, P., Jeong, T., Xu, L., 2023a. Transformer Based Water Level Prediction in Poyang Lake, China. Water 15, 576. https://doi.org/10.3390/w15030576.
Xu, Y., Lin, K., Hu, C., Wang, S., Wu, Q., Zhang, L., Ran, G., 2023b. Deep transfer learning based on transformer for flood forecasting in data-sparse basins. J. Hydrol. 625, 129956 https://doi.org/10.1016/j.jhydrol.2023.129956.
Xu, D., Zhang, Q., Ding, Y., Zhang, D., 2022. Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. Environ. Sci. Pollut. Res. 29, 4128–4144. https://doi.org/10.1007/s11356-021-15325-z.
Yan, H., DeChant, C.M., Moradkhani, H., 2015. Improving Soil Moisture Profile Prediction With the Particle Filter-Markov Chain Monte Carlo Method. IEEE Trans. Geosci. Remote Sens. 53, 6134–6147. https://doi.org/10.1109/TGRS.2015.2432067.
Yang, R., Luo, F., Ren, F., Huang, W., Li, Q., Du, K., Yuan, D., 2022. Identifying Urban Wetlands through Remote Sensing Scene Classification Using Deep Learning: A Case Study of Shenzhen. China. ISPRS Int. J. Geo-Inf. 11, 131. https://doi.org/10.3390/ijgi11020131.
Yang, T., Sun, F., Gentine, P., Liu, W., Wang, H., Yin, J., Du, M., Liu, C., 2019b. Evaluation and machine learning improvement of global hydrological model-based flood simulations. Environ. Res. Lett. 14, 114027 https://doi.org/10.1088/1748-9326/ab4d5e.
Yang, S., Yang, D., Chen, J., Zhao, B., 2019a. Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model. J. Hydrol. 579, 124229 https://doi.org/10.1016/j.jhydrol.2019.124229.
Yao, Y., Rosasco, L., Caponnetto, A., 2007. On Early Stopping in Gradient Descent Learning. Constr. Approx. 26, 289–315. https://doi.org/10.1007/s00365-006-0663-2.
Yaseen, Z.M., Sulaiman, S.O., Deo, R.C., Chau, K.-W., 2019. An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. J. Hydrol. 569, 387–408. https://doi.org/10.1016/j.jhydrol.2018.11.069.
Yeh, W.-W.-G., 1986. Review of Parameter Identification Procedures in Groundwater Hydrology: The Inverse Problem. Water Resour. Res. 22, 95–108. https://doi.org/10.1029/WR022i002p00095.
Yu, S., Ma, J., 2021. Deep Learning for Geophysics: Current and Future Trends. Rev. Geophys. 59, e2021RG000742 https://doi.org/10.1029/2021RG000742.
Yu, X., Shen, J., Du, J., 2020. A Machine-Learning-Based Model for Water Quality in Coastal Waters, Taking Dissolved Oxygen and Hypoxia in Chesapeake Bay as an Example. Water Resour. Res. 56, e2020WR027227 https://doi.org/10.1029/2020WR027227.
Yu, P.-S., Yang, T.-C., Chen, S.-Y., Kuo, C.-M., Tseng, H.-W., 2017. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. J. Hydrol. 552, 92–104. https://doi.org/10.1016/j.jhydrol.2017.06.020.
Zaherpour, J., Mount, N., Gosling, S.N., Dankers, R., Eisner, S., Gerten, D., Liu, X., Masaki, Y., Müller Schmied, H., Tang, Q., Wada, Y., 2019. Exploring the value of machine learning for weighted multi-model combination of an ensemble of global hydrological models. Environ. Model. Softw. 114, 112–128. https://doi.org/10.1016/j.envsoft.2019.01.003.
Zhang, C., Li, G., Du, S., Tan, W., Gao, F., 2019. Three-dimensional densely connected convolutional network for hyperspectral remote sensing image classification. J. Appl. Remote Sens. 13, 016519 https://doi.org/10.1117/1.JRS.13.016519.
Zhang, C., Harrison, P.A., Pan, X., Li, H., Sargent, I., Atkinson, P.M., 2020. Scale Sequence Joint Deep Learning (SS-JDL) for land use and land cover classification. Remote Sens. Environ. 237, 111593 https://doi.org/10.1016/j.rse.2019.111593.
Zhang, C., Brodeur, Z.P., Steinschneider, S., Herman, J.D., 2022. Leveraging Spatial Patterns in Precipitation Forecasts Using Deep Learning to Support Regional Water Management. Water Resour. Res. 58, e2021WR031910 https://doi.org/10.1029/2021WR031910.
Zhi, W., Feng, D., Tsai, W.-P., Sterle, G., Harpold, A., Shen, C., Li, L., 2021. From Hydrometeorology to River Water Quality: Can a Deep Learning Model Predict Dissolved Oxygen at the Continental Scale? Environ. Sci. Technol. 55, 2357–2368. https://doi.org/10.1021/acs.est.0c06783.
Zhong, Z., Sun, A.Y., Yang, Q., Ouyang, Q., 2019. A deep learning approach to anomaly detection in geological carbon sequestration sites using pressure measurements. J. Hydrol. 573, 885–894. https://doi.org/10.1016/j.jhydrol.2019.04.015.
Zhou, Z., Shi, L., Zha, Y., 2020. Seeing macro-dispersivity from hydraulic conductivity field with convolutional neural network. Adv. Water Resour. 138, 103545 https://doi.org/10.1016/j.advwatres.2020.103545.
Zhou, K., Zheng, Y., Li, B., Dong, W., Zhang, X., 2019. Forecasting Different Types of Convective Weather: A Deep Learning Approach. J. Meteorol. Res. 33, 797–809. https://doi.org/10.1007/s13351-019-8162-6.

Glossary

AE: Autoencoder
AI: Artificial Intelligence
ANN: Artificial Neural Network
CNN: Convolutional Neural Network
DAE: Denoising Autoencoder
DL: Deep Learning
FCNN: Fully Connected Neural Network
GAN: Generative Adversarial Network
GPU: Graphical Processing Unit
GRU: Gated Recurrent Unit
IOT: Internet of Things
LSTM: Long Short-Term Memory
ML: Machine Learning
MLP: Multi-Layer Perceptron
PCA: Principal Component Analysis
PGDL: Physics Guided Deep Learning
ResNet: Residual Neural Network
RF: Random Forest
RNN: Recurrent Neural Network
SOM: Self Organizing Maps
SVM: Support Vector Machines
TL: Transfer Learning
VAE: Variational Autoencoders
WRF: Weather Research Forecast
XAI: eXplanable Artificial Intelligence
XGBoost: Extreme Gradient Boosting